8887-EN-01

DTIC

APPLICATION OF GEOSTATISTICAL METHODS AND WAVELETS TO THE ANALYSIS OF HYPERSPECTRAL IMAGERY AND THE TESTING OF A MOVING VARIOGRAM

Second Interim Report (RSSUSA - **5/2**)

Dr Margaret A. Oliver

August 2000 to October 2000

United States Army

ENVIRONMENTAL RESEARCH OFFICE OF THE U.S. ARMY

London, England

CONTRACT NUMBER - N68171-00-M-5508

Contractor - Approved for Public Release; distribution unlimited

20010508 096

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	27. 11.00	Interim August 2000 – October 2000

4. TITLE AND SUBTITLE
Application of geostatistical methods and wavelets to the analysis of hyperspectral imagery and the testing of a moving variogram

5. FUNDING NUMBERS
N68171
00-M-5508

6. AUTHOR(S)
Dr Margaret A Oliver

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Reading, Whiteknights,
Reading, RG6 2AH, UK

8. PERFORMING ORGANIZATION REPORT NUMBER
RSS USA-5/2

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
USARDSG-UK, Environmental Sciences Branch
Edison House, 233 Old Marylebone Road,
London, NW1 5TH, UK

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Interim Report:
Summary of work to date

12a. DISTRIBUTION/AVAILABILITY STATEMENT
No limitation on distribution/availability

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The work in this report focuses on an analysis of the National Soil Inventory of England and Wales. The aim was to compare geostatistical methods, mainly ordinary kriging and factorial kriging and wavelet analysis, on a different kind of data from imagery. The data were from sampling locations on a 5-km grid.To provide an area as close to a square as possible for the wavelet analysis, just over 3000 points were selected from the total of over 5000. Two variables were selected for analysis, pH and zinc. The variogram of pH showed that there was long-range trend in the data which meant that this had to be removed for the geostatistical analysis. Trend makes the geostatistical analysis more complex, whereas the wavelet analysis is not affected by it. Zinc was markedly skewed and the data were transformed to common logarithms for the geostatistical analysis, which again was not necessary for the wavelet analysis. The results have shown some interesting features. There appears to be no local non-stationarity in these data, which meant that kriging performed better than the wavelet analysis in terms of the distribution of the errors for the 10-km subsample. However, for the 40-km subsample the wavelet analysis performed better. The variograms for both properties were nested and the short-range variation was evident in the high frequency wavelet transform for the 20-km grid. The variogram can provide a guide as to what sampling interval should be focused on in a multiresolution analysis using wavelets.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Keywords: factorial kriging, trend, wavelets, multiresolution, soil	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

ATTACHMENT 1
PAGE 1 OF 2

A geostatistical and wavelet analysis of the National Soil Inventory of England and Wales

Introduction

It was decided while on a visit to TEC by Mr W. Clark that Mr E. Bosch and Dr M. A. Oliver should extend the comparison of the geostatistical and wavelet analysis that they had done on the Fort A. P. Hill SPOT image to a large data set of soil information. The reason for this was to see how the techniques performed when the initial data are a sample rather than complete cover as in the image. For the SPOT image we had full cover of pixel information which we then sampled. Kriging and the low frequency wavelet coefficients were used to restore the data that had been removed (see report ?? and Oliver *et al.*, 2000). For the soil data we started with sample information and resampled this for the analyses. The wavelet analysis of the soil data was done by Mr E. Bosch during Dr Oliver's visit to TEC in June 2000.

The Soil Data

The soil data that we have analysed are part of the National Soil Inventory (NSI) of England Wales, which was carried out by the Soil Survey of England and Wales between 1978 and 1983 (McGrath & Loveland, 1992). The aim of the survey was to provide a record of the soil information in these countries and both toxicity and deficiency of some elements of the soil that affect both grazing animals and arable crops at the national level. For the NSI to be an unbiased estimate of the distribution of types of land and their properties, strict protocols were applied to site location and description, soil sampling strategy, and soil profile description. This was very unlike the practice of 'free' soil survey which is commonly used to produce conventional soil maps (Avery, 1987). Considerable effort also went into quality control of pre-treatment and analysis of the samples, data recording, error trapping and construction of the database, because of the number of samples and the magnitude of the subsequent analytical programme (Loveland, 1990; McGrath & Loveland, 1992).

The number of samples was restricted to those falling at the intersectoins of a 5-km orthogonal grid. The sampling grid was offset 1 km north and east of the origin of the Ordnance Survey National Grid. If the sampling point fell on anything other than land, e.g. on a road, building, water-body etc., then the sampling point was moved 100 m north of the grid node. If that failed to locate suitable soil, then the point was moved 100 m west *from the originally intended point.* This process was repeated in steps of 100 m and 200 m from the grid node, in the order north, east, south and west. If no suitable soil was found after this procedure, then the site was abandoned for sampling purposes, although the land-use at the original sampling point was recorded so that the inventory was complete and to make clear the reason for the deviation. If a new sampling point was found, then the standard procedure for description and sampling was followed at that point (see below). In this way, an unbiased record of the occurrence of various forms of land-use was maintained.

The principal interest was in agricultural land. No attempt was made to devise a sampling strategy to cover urban areas adequately. In total 5691 sites were sampled. The grid-reference located the site to within 10m on the ground, i.e. to an accuracy which would place any return visit within the original soil sampling sub-grid (see below).

Sampling

The soil profile was described in a pit dug to 80 cm (or less if rock was encountered) at each sampling point, using standard terminology (Hodgson, 1974). However, sampling was restricted to the uppermost 15 cm of *mineral* soil (or less if rock intervened), or of peat, as appropriate, i.e. litter layers were not sampled, as they were regarded as ephemeral. The actual sampling depth was recorded. Twenty-five cores of soil were taken at the nodes of a 4m grid within a 20 m x 20 m square centred on the Ordnance Survey (OS) 5-km grid-point. The cores were taken with a screw-type, mild-steel auger, to avoid contamination from traces of elements such as chromium and manganese present in stainless, plated or similar special steels. The cores of soil were bulked and mixed well in the field and double-bagged. in food-grade polythene bags, and a waterproof and rot-proof label ('Synteape') placed between the bags.

Samples were air-dried and milled in a mild-steel roller-mill (Waters & Sweetman, 1955) to pass a 2-mm aperture sieve. Preliminary work had shown that no detectable contamination of the samples arose from this procedure. The resulting data set comprises up to 127 analytical and descriptive parameters for each of 5691 points across England and Wales (Loveland, 1990; McGrath & Loveland, 1992). This collection of data is a unique and invaluable resource,

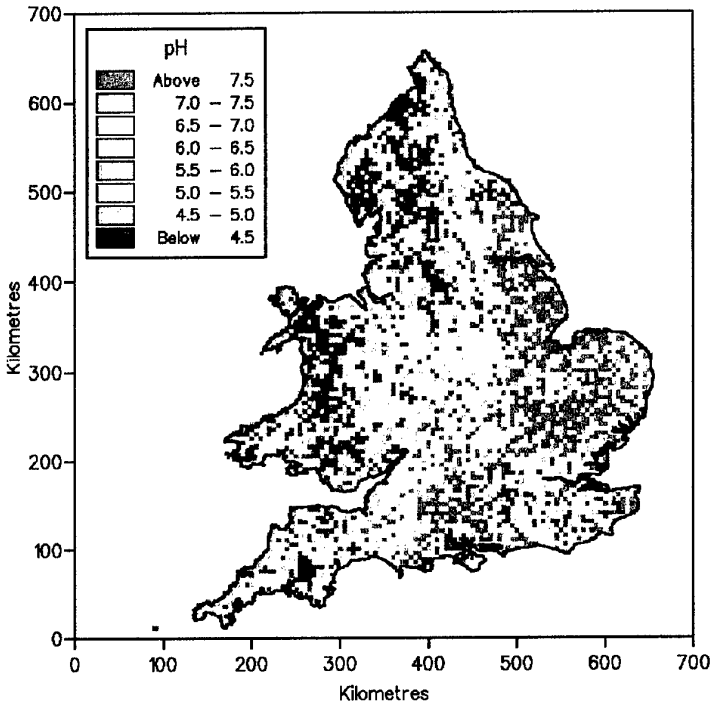
In this analysis we have examined only pH and total Zinc because they represent other variables well. From a principal components analysis (PCA) Zn was seen to load heavily on the first component and pH on the third axis, which reflected the effects of pareent material (PM) and leaching. The pH was measured by a combination electrode and pH meter in a 1:2.5 soil-water suspension (MAFF, 1986) on soil <2-mm. Zinc was determined using the <150 micrometre soil. It was extracted by *aqua regia*, and determined by ICP-AES (RES) (McGrath & Cunliffe, 1985).

Analysis

Figures 1a and 2 a show the full set of pixel information on the 5-km grid for pH and Zn. (They have a different colour scale from the remaining maps because they were prepared on a different computer for the Ministry of Agriculture analysis. Nevertheless the variation can be compared and the relative area that we have worked on). The analysis for this project was carried out on a subset of the full NSI data. This was because the wavelet analysis requires a set of data that is square and can be sampled in octaves. The outline of the data for England and Wales is irregular and we selected data from the central part of the country to obtain as large a square as possible that would suit the needs of the analysis. This resulted in a data set with 3500 sites. For the wavelet analysis the data were 'padded' with zeros so that there were no gaps. The latter arose because of the shape of the coastline and the urban areas within the country that were not sampled (see Figures 1a and 2a); they appear as white patches. Figures 1b and 2b show the data that were extracted for the analysis in this report.

Figure 1. Raw data for pH

a) Original data for pH



b) Original data for pH selected from the full data for analysis

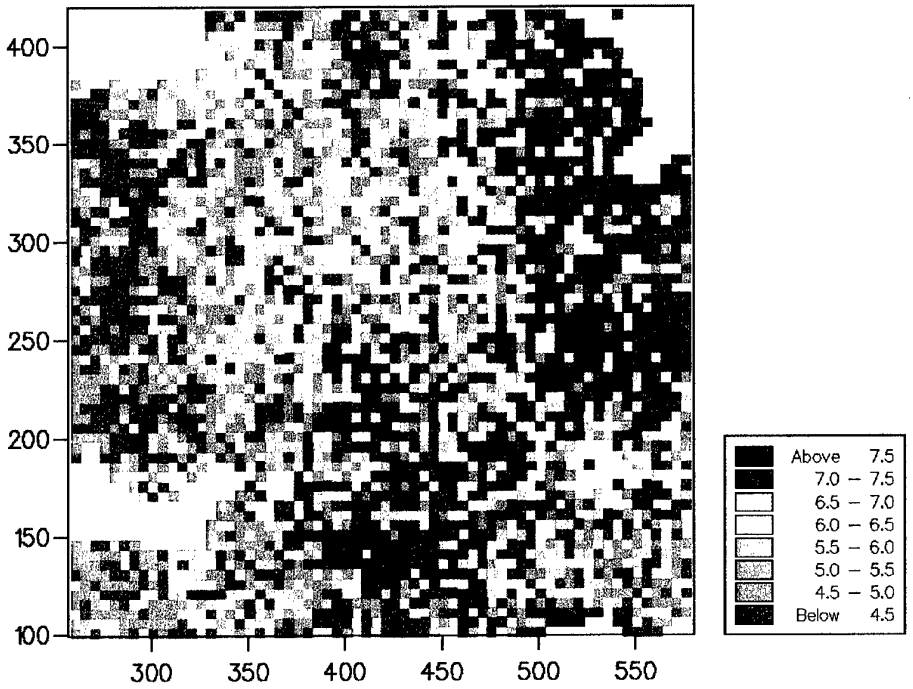
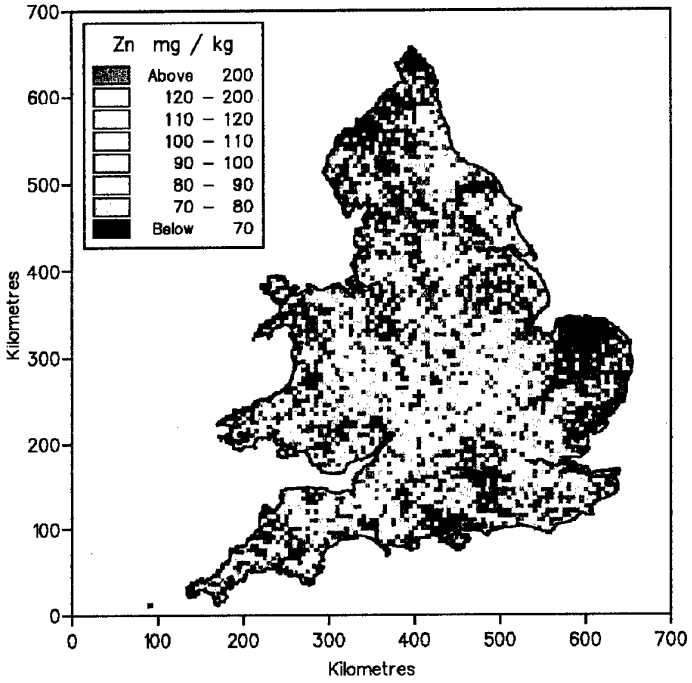


Figure 2. Raw data for Zinc.

a) Original data for Zn



b) Original data for Zn selected from the full data for analysis

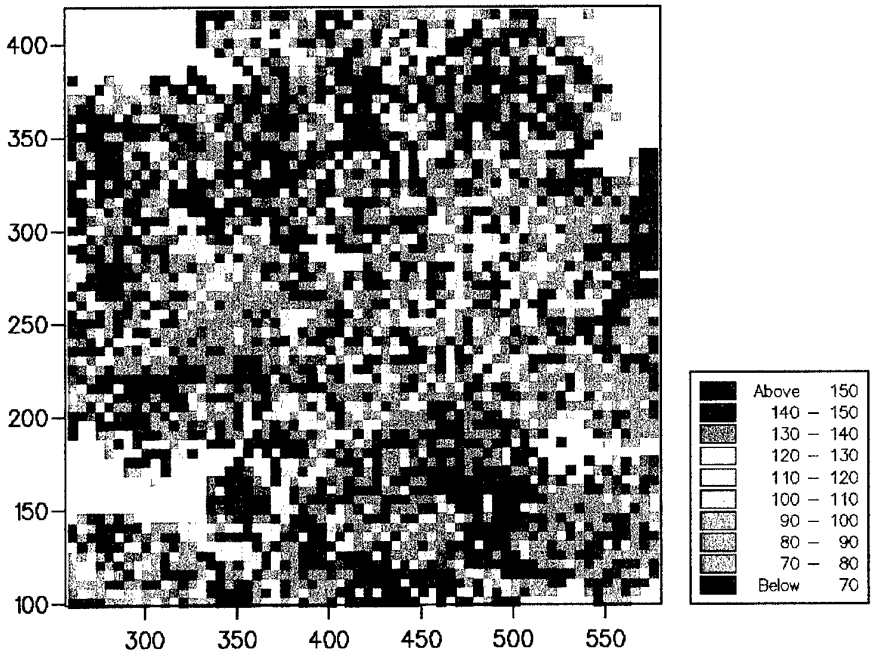
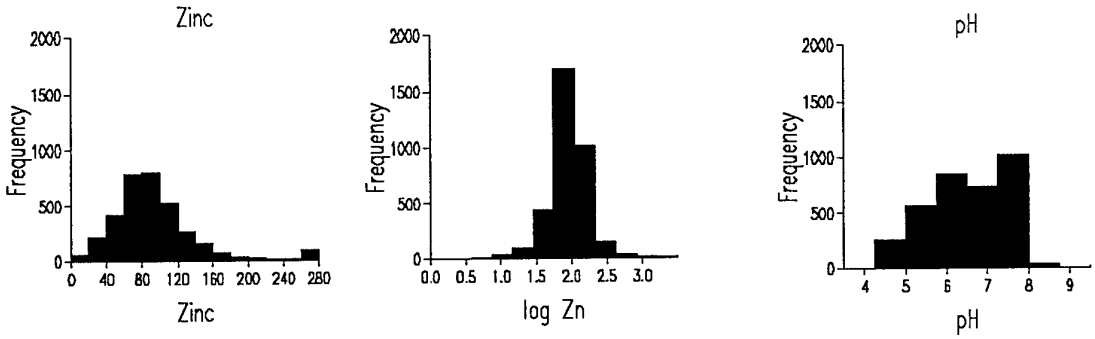


Table 1 gives the summary statistics of the subset of the data. Zinc was strongly positively skewed which can be seen from Table 1 and Figure 3 a, therefore it was transformed to common logarithms (log_{10}), Figure 3 b. This transformation has produced a log-normal distribution which is common for many elements, Figure 3 b and Table 1. The histogram for pH shows that this has a close to normal distribution, Table 1. A near-normal distribution is necessary for the variogram analysis because it is based in variances, which are unstable if the data do not have a near normal distribution.

Figure 3. Histograms from the subset of the NSI data for Zinc and pH.

a) Zinc b) $Log_{10}Zn$ c) pH



The data on the 5-km grid were further subsampled to compare the results of data reconstruction by both kriging and wavelet analysis. The subsampling produced grids of 10-km (1 site in every block of 4 sites resulting in 869 sites), 20-km (1site in every block of 16 sites resulting in 219 sites), and 40-km (1 site in every block of 64 sites resulting in 57 sites).

Table 1. Summary statistics for pH and Zinc for the subset of data used in the analysis.

Statistic	pH	Zinc	$Log_{10}Zn$
Minimum	3.200	6.00	0.7782
Maximum	9.100	3648.0	3.562
Mean	6.094	103.95	1.934
Variance	1.627	14273.2	0.0595
Standard deviation	1.275	119.47	0.2440
Skewness	-0.2377	14.03	-0.0002

GEOSTATISTICAL ANALYSIS

Variogram analysis

The spatial structure in the data was determined by computing the experimental variograms of pH and $\log_{10}$Zn. For the full set of data Zn and pH showed no marked evidence of anisotropy, therefore omnidirectional variograms only were computed. For both the full data and the subset the experimental variogram of pH showed evidence of trend. The semivariances continued to increase after an initial sill had been reached (Figure 4 a). This suggests the presence of smooth continuous variation that violates the assumptions of geostatistics, which assumes that the variable is random. Therefore, we modelled the trend by linear and quadratic functions of the co-ordinates so that the analysis could be done on the residuals from the trend. The linear function was less effective in accounting for the trend than the quadratic one: the latter removed over 30% of the trend in both cases. The variogram was then computed afresh on the residuals, and this now shows a more simple bounded form, Figure 4 b.

Most of the variograms were fitted by nested functions. The models fitted to the data included single exponential, spherical, and power functions including linear, double exponential and spherical, and exponential with linear functions. For $\log_{10}$Zn (Figure 4 c) and pH of the residuals double spherical models provided the best fit. The equations for the models are given below.

Double exponential

$$\gamma(h) = c_0 + c_1\{1 - \exp(-h/r_1)\} + c_2\{1 - \exp(-h/r_2)\}$$

where c_1 and r_1 are the sill and distance parameter of the first structure, and c_2 and r_2 are the sill and distance parameter of the second structure.

Double spherical

$$\gamma(h) = c_0 + c_1\left\{\frac{3h}{2a_1} - \frac{1}{2}\left(\frac{h}{a_1}\right)^3\right\} + c_2\left\{\frac{3h}{2a_2} - \frac{1}{2}\left(\frac{h}{a_2}\right)^3\right\} \qquad \text{for } h \le a_1$$

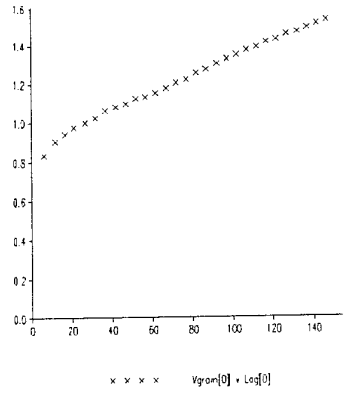
$$\gamma(h) = c_0 + c_2\left\{\frac{3h}{2a_2} - \frac{1}{2}\left(\frac{h}{a_2}\right)^3\right\} \qquad \text{for } a_1 \le h \le a_2$$

$$\gamma(h) = c_1 + c_2 \qquad \text{for } h \ge a_2$$

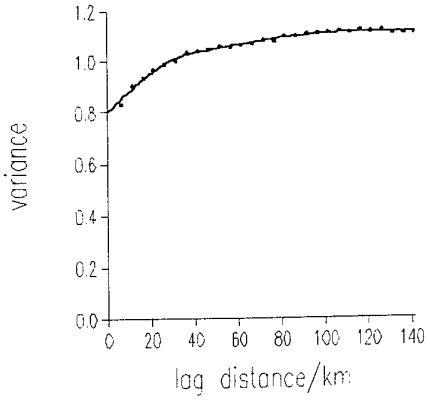
where c_1 and a_1 are the sill and distance parameter of the first structure, and c_2 and a_2 are the sill and distance parameter of the second structure.

Figure 4. Experimental variograms (symbols) and fitted models (lines): a) raw values of pH, b) pH residuals, and c) $\log_{10}$Zn.

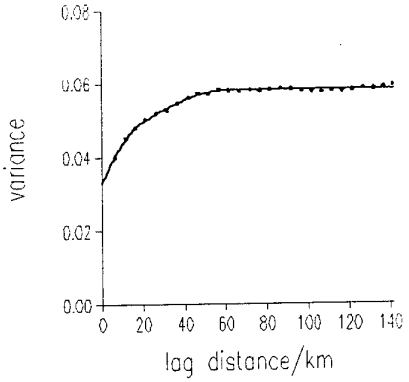
a) pH – raw data



b) pH residuals



c) Log_{10}Zn



These results show that there are two main scales of spatial variation: a short-range component of about 18 km for $\log_{10}$Zn and 37 km for pH (residuals), and a long-range component of 61 km for $\log_{10}$Zn and 118 km for pH (residuals). The average short-range component for the full data for the range of properties examined was 24 km, and the average of the long-range component was 89 km. There are slight differences in the ranges for these subsets, but they are within similar orders of magnitude. A characteristic of the variograms of the subset and of the full data is their large nugget variance (c_0): it is more than 60% of the sill variance for most variables. Most of the nugget variance can be accounted for by variation over distances less than the sampling interval of the grid. This shows that the 5-km grid interval misses a considerable proportion of the variation in the soil.

Figure 1 shows the pixel map of pH. There are two spatial scales of variation evident in the map. Areas with a pH of less than 6 are in the western part of the country in the main, which is also where the main uplands are, and where agriculture is dominated by grassland systems - optimum pH between 5 and 6.. These are also the wettest parts of England and Wales. Much of central and eastern England has pH values of 7 and above, partly reflecting geology and the distribution of calcareous soils, but also the widespread use of lime on arable soils (optimum pH *c.* 6.5 - 7.5). There are areas of lower pH in the S associated with the Tertiary sands and gravels. The E-W differences in pH values reflect the pattern of rainfall as well as elevation and land-use. Figure 2 shows the original values as a pixel map for total Zn. There are many areas with large concentrations, and the most extensive of these follows the Jurassic clay band from SW to NE across the country. There are other areas trending N to S from the Midlands of England to Tynemouth (not on the subset map). These seem to be associated with the Carboniferous shales and sandstones, as do the areas of large concentrations in central Wales.

Kriging

Ordinary kriging and factorial kriging (kriging analysis) have been described in earlier reports (Contract Nos. N68171-97-C-9029; N68171-98-M-5311). Ordinary kriging was used to reconstruct the data after subsampling them to produce smaller data sets. Punctual

kriging was used so that the estimates and maps could be compared with predictions from subsets of the data. The estimation grid was chosen to coincide with the 5-km sampling grid. Estimates were made at the nodes of this grid so that we could compare the kriged estimates at the sampling points with the original values where these had been removed. At the places where there were data punctual kriging returns the sample value there. The original variograms were used for the analysis because it is unlikely that their structure would change over time. In addition those from the subsets have large nugget variances and they are less reliable because there are few comparisons for each semivariance, especially for the 40-km grid.

For pH ordinary kriging was done on the residuals and the quadratic trend added back to the estimated residuals afterwards.

The ordinary kriged logarithmically transformed predictions for Zinc were back-transformed for mapping so that the variation could be seen on the original scale in which the variable had been measured. This is not straightforward because the kriging variance must be taken into account. The equation for back-transformation is:

$$\hat{Z} = \exp\{\hat{Y}(\mathbf{x}_0) \times \ln 10 + 0.5\sigma_Y^2(\mathbf{x}_0) \times (\ln 10)^2\}$$

where $\hat{Y}(\mathbf{x}_0)$ is the estimated value of $\log_{10}$Zn at $\mathbf{x}_0$ and σ^2_Y is the estimation variance.

Factorial kriging was done on the full set of data to examine the different scales of variation in the data and to compare the results with those of the multi-resolution wavelet analysis. The aim is to filter out the different scales of variation, so that the independent components of the spatial structure can be examined as an aid to further interpretation. Factorial kriging estimates the long- and short-range components separately. The variation is nested for both pH and Zn; the variograms have two spatial structures. The pixel maps of the raw data, Figures 1 and 2 suggest that there are two scales of variation, and this is confirmed by the variogram results. The variation at the longer scale appears to be related to the geology for Zn and pH and also rainfall and elevation for the latter.

Zinc is a good example of many of the other variables for this analysis. For the long-range

WAVELET ANALYSIS

The method for this analysis was described in a previous report (N68171-98-M-5311) and also in Oliver *et al.* (2000) for SPOT image data. This is the first analysis that we are aware of using soil data in two dimensions. There are few data sets in the world for soil that are on a grid and would provide adequate data for this analysis. Wavelets enable data reconstruction and multi-resolution analysis by deriving the low frequency and high frequency coefficients from the data. The low frequency wavelet transform has been used to restore the data from the subsamples on the original 5-km grid and to identify the long-range spatial component at the coarser resolutions. The average of the high frequency wavelet transforms has been used to identify the short-range component. The advantage of wavelet analysis at the outset for the pH data is that there is no need to take account of trend. An important advantage of this analysis is that it is unaffected by non-stationarity.

RESULTS FOR pH

The following series of maps (Figures 5 to 7) shows the reconstructed values of pH from ordinary kriging and the low frequency wavelet coefficients for the three sampling grids. One noticeable difference between the maps is that the kriged maps appear more 'spotty'. This is because kriging returns the sample value at the data point, whereas the wavelet analysis is a predicted value at the data points as well as at other points. Another difference arises from the fact that the data were padded for the wavelet analysis with zeros – these are the larger blue areas beyond the coastline and also the urban areas where there were no sampling locations. Figure 5 b for the data on a 10-km grid there is slightly more of the original detail in the variation evident, whereas the kriged map (Figure 5 a) shows the effect of smoothing from kriging.

For Figure 6 a and b the effects of the greatly reduced number of sampling sites is evident in the loss of detail. For Figure 6 a the margin around the map is because there were no data there to krige from. Kriging, Figure 6 a, has smoothed the variation more than wavelet analysis, Figure 6 b, and the spotty appearance of the former map is the effect of punctual kriging. Figure 7 a and b shows the kriged and low frequency wavelet coefficients for the 40-km grid. It is clear that much detail has been lost and that there is more difference between these two maps than between those in Figures 5 and 6. Visually the wavelet analysis appears to have performed better at this level of sampling which is what we found for the image data (Oliver *et al.*, 2000). The more sparse the sampling the better the wavelet analysis appears to perform in comparison with kriging.

Figures 8 to 10 show the maps of the comparisons between the predictions from ordinary punctual kriging and the low frequency wavelet transform, and the original values at the sampling sites of the 5-km grid. Figure 8 a and b show the comparisons, i.e. the absolute differences, for predictions based on the 10-km sampling grid. Figure 8 a for the kriged comparisons is a more spotty map than the one from the wavelet analysis: the sampling points are evident as the blue pixels where there is no error. For the wavelet analysis for this sampling grid there are fewer zero or small errors than for kriging, This is confirmed by the histograms of the differences, Figure 11 a and b. Kriging also appears to perform better for predicting the values from the 20-km grid than the wavelet analysis in terms of the small errors, Figure 9 a and b. The histogram of the kriged differences, Figure 11 c, is somewhat misleading because there are fewer comparisons for kriging than for the wavelet analysis, and it is likely that there would be more of the larger errors than is evident in the histogram. The slight negative skewness in this histogram suggests that there is some bias in the predictions. The histograms, Figure 11 c and d, for this sampling grid (20-km) are more similar than for the 10-km one. The comparisons for the predictions from the 40-km grid suggest that the wavelet analysis has performed slightly better at this level of sampling, which was the case for the SPOT image data (Oliver *et al.*, 2000). These histograms, Figure 11 d and e show that the wavelet analysis has more smaller errors. There are fewer comparisons for kriging because the method requires a

minimum of 4 points within the search radius and this fails at the margins of the error when the sampling points become sparse.

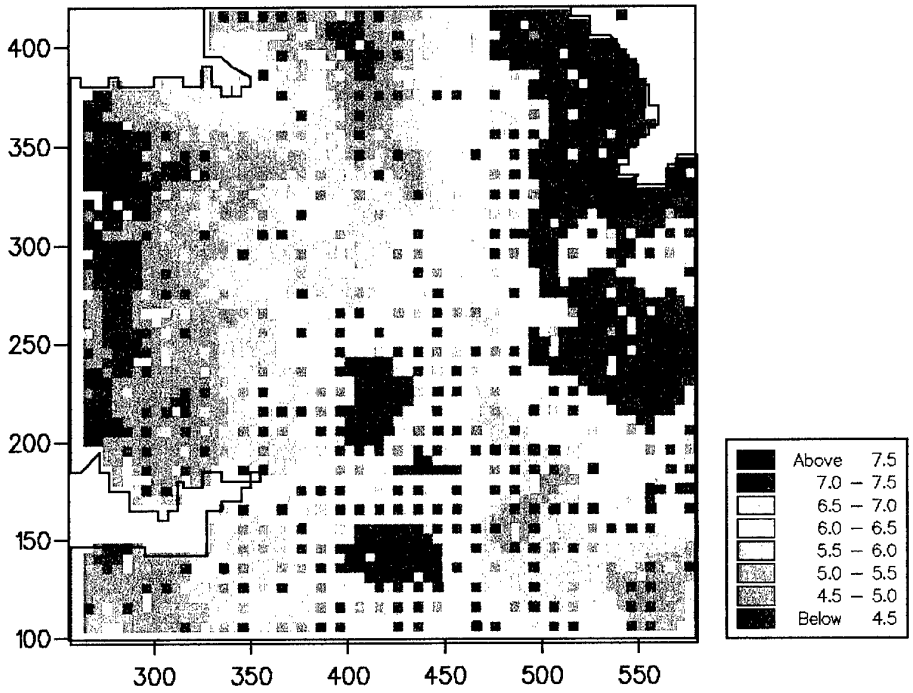
Summary

These results are interesting when compared with the analysis of the SPOT data. The NSI data appear not contain locally non-stationary data for pH. These would occur where there are marked boundaries in the soil, for example. At the sampling interval used here of 5-km local non-stationarity is less likely than for more intensively sampled data and remotely sensed data. Therefore, the errors for kriging are less than they were for the analysis of the SPOT data where there were marked changes at lakes and other boundaries causing local non-stationarity. Since kriging is an exact interpolator and wavelet analysis is not, there remains the need to combine the methods. It seems that some progress on this has been made at the Centre de Géostatististique, Fontainebleau. However, at the moment it is difficult to ascertain the extent of this and we shall endeavour to take this forward.

Another point of interest from this analysis is that the pH data form the NSI survey contain long distance trend. This means that part of the variation depends on the spatial coordinates. This violates the assumptions of geostatistics in the same way as local trend or drift, i.e. local non-stationarity. This affected the variogram, as was evident above, Figure 4 a and b, and meant that we had to remove the trend and do the analysis on the residuals, and add back the trend after kriging. This is clearly a considerable amount of additional effort over and above the straightforward analysis. It is evident from the results of the wavelet analysis that the prediction are unaffected by the trend. Therefore, if there is a choice of method available – situations with known trend present would benefit from the wavelet analysis.

Figure 5. Predictions of pH at a 5-km interval from data on a 10-km grid

a) pH - kriged estimates form the data on a 10-km grid



b) pH - low frequency wavelet transform for data on a 10-km grid

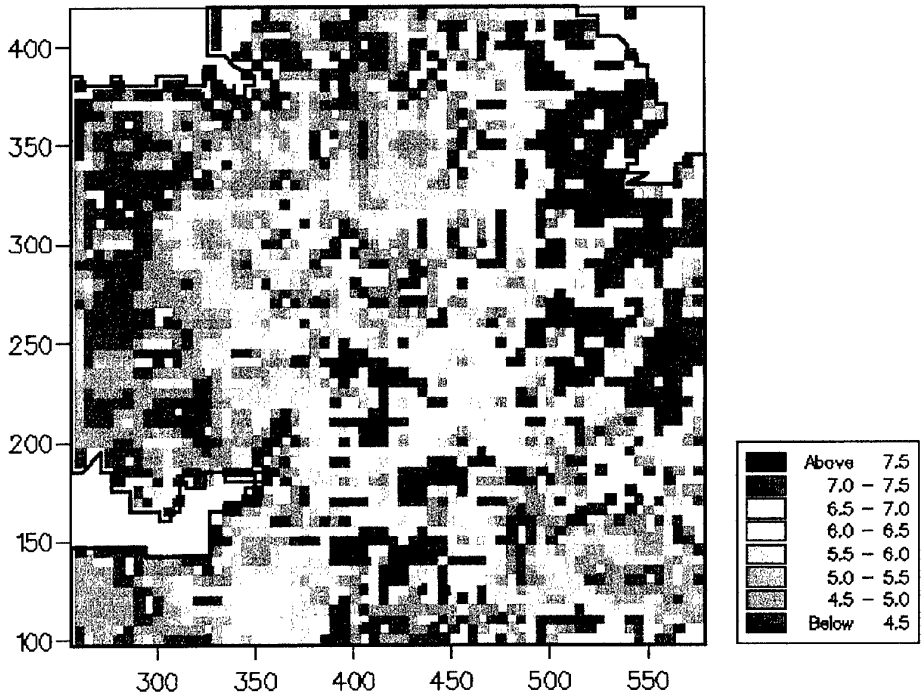
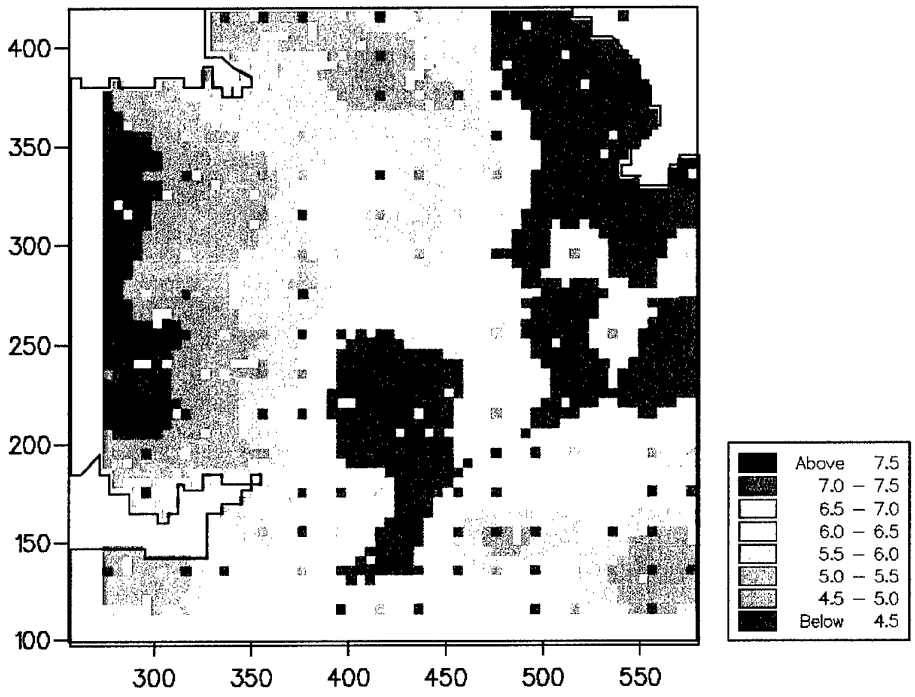


Figure 6. Predictions at a 5-km interval from data on a 20-km grid

a) pH – kriged estimates from data on a 20-km grid



b) pH- low frequency wavelet transform for data on a 20-km grid

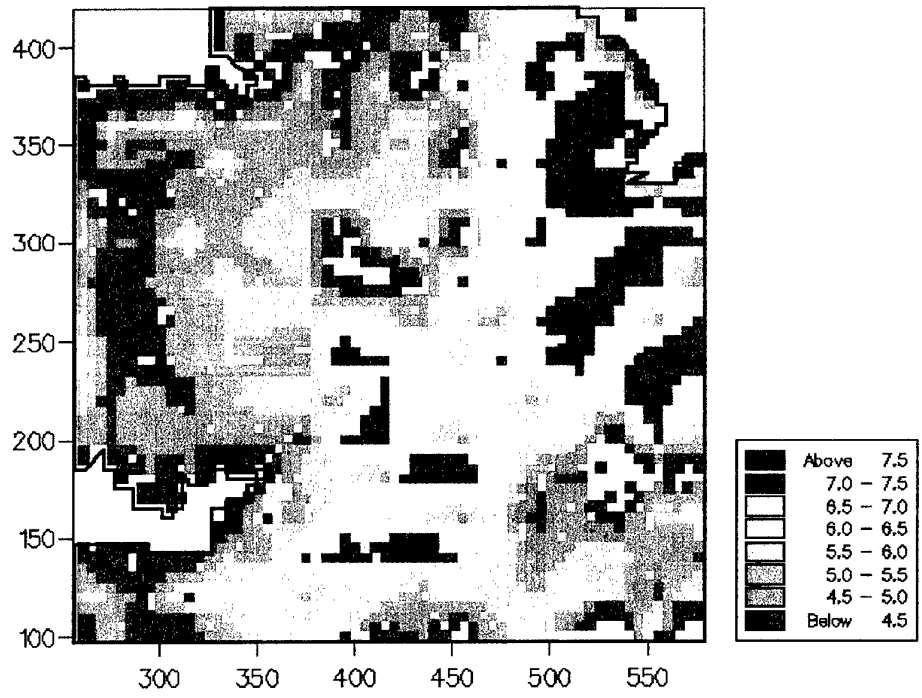
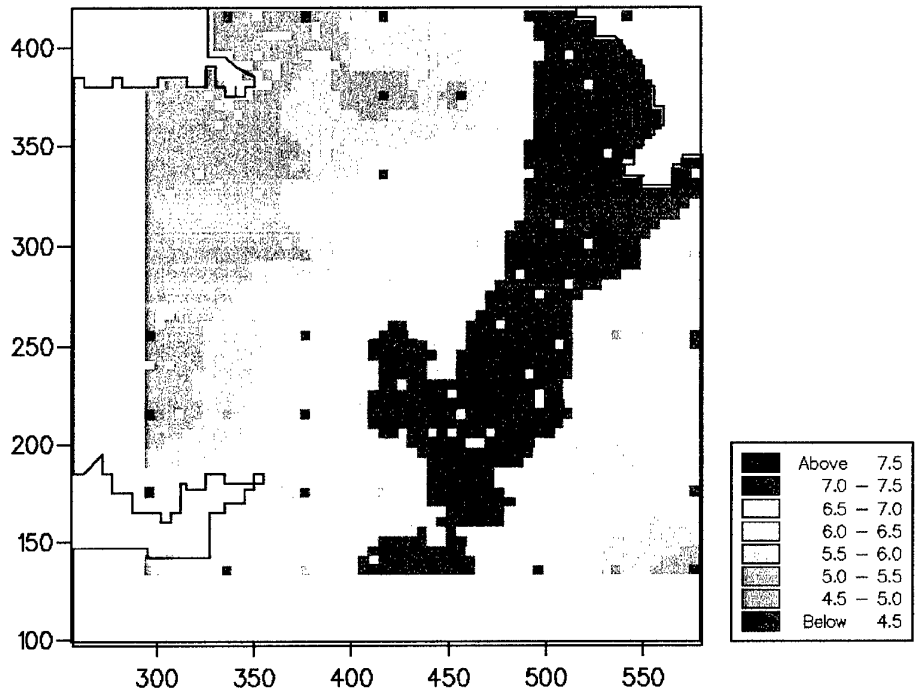


Figure 7. Predictions at a 5-km interval from data on a 40-km grid.

a) pH – kriged estimates from data on a 40-km grid



a) pH – low frequency wavelet transform for data on a 40-km grid

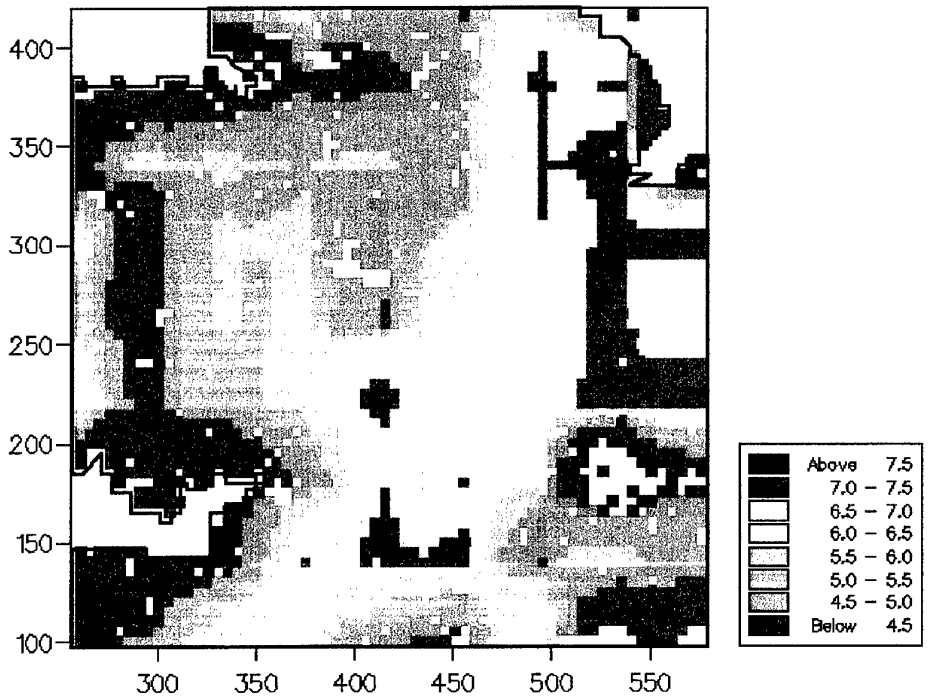
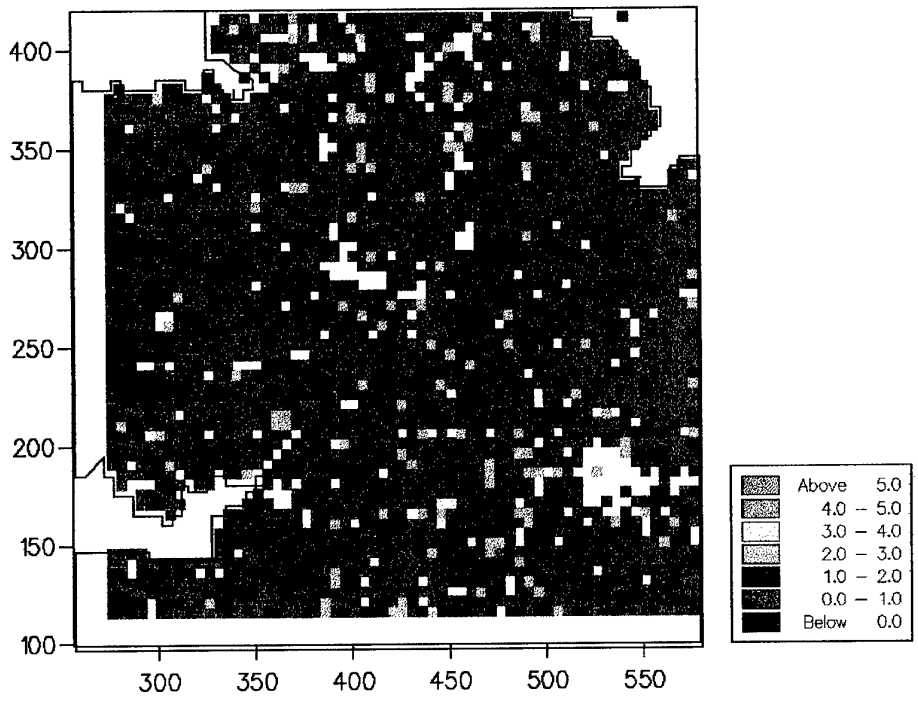


Figure 8. Comparisons between estimates from data on a 10-km grid

a) pH - comparisons for kriging



b) pH – comparisons for wavelet analysis

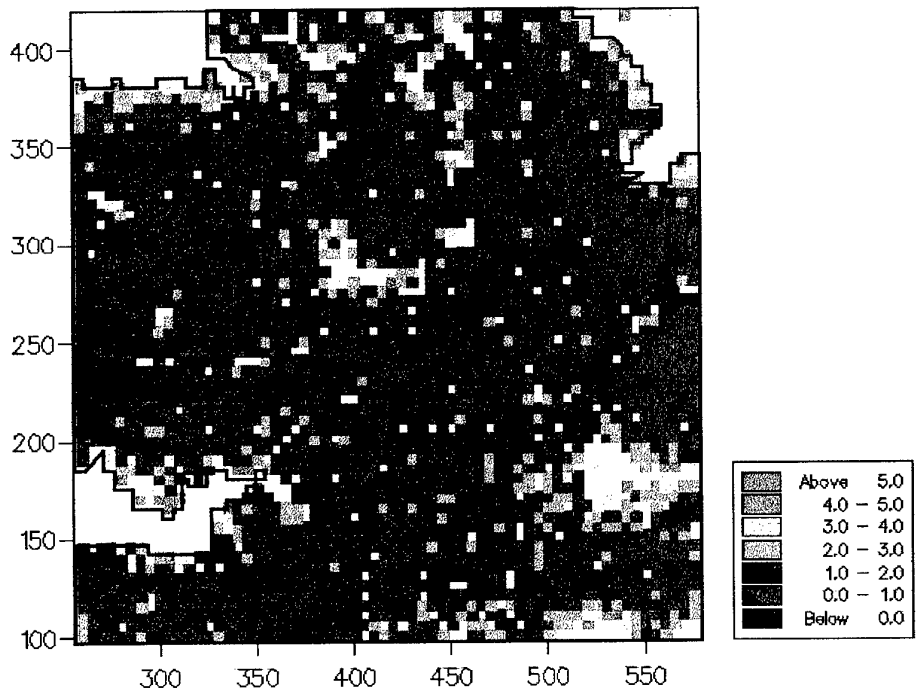
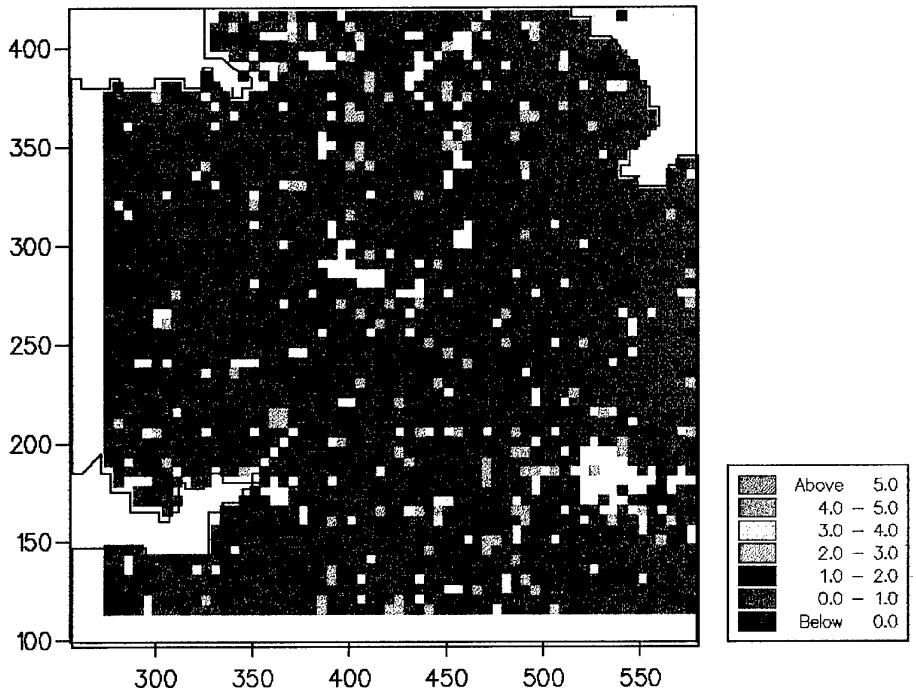


Figure 9. Comparisons for estimates from data on a 20-km grid

a) pH – comparisons for kriging



b) pH – comparisons for wavelet analysis

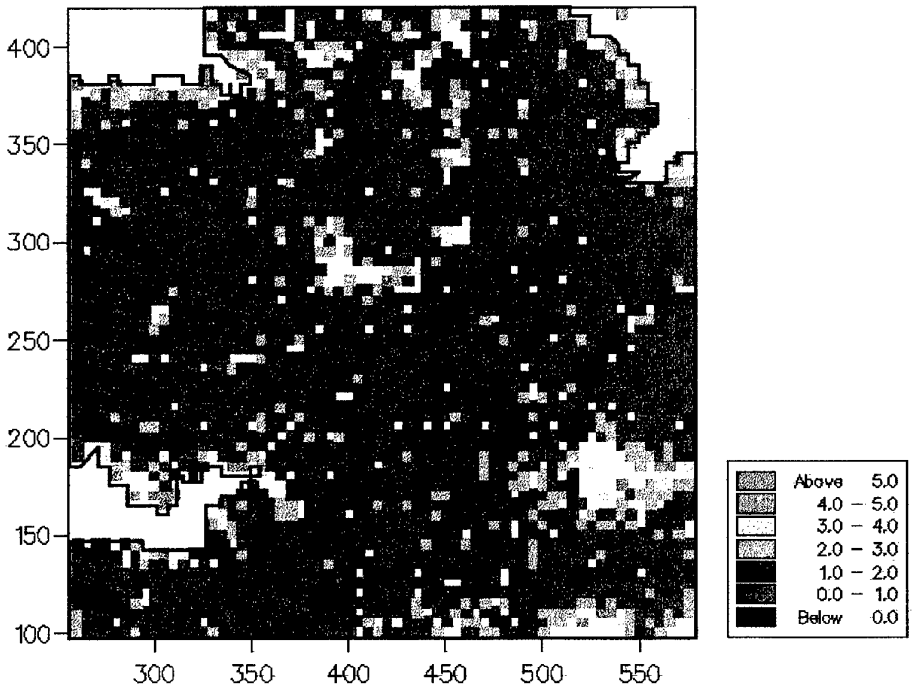
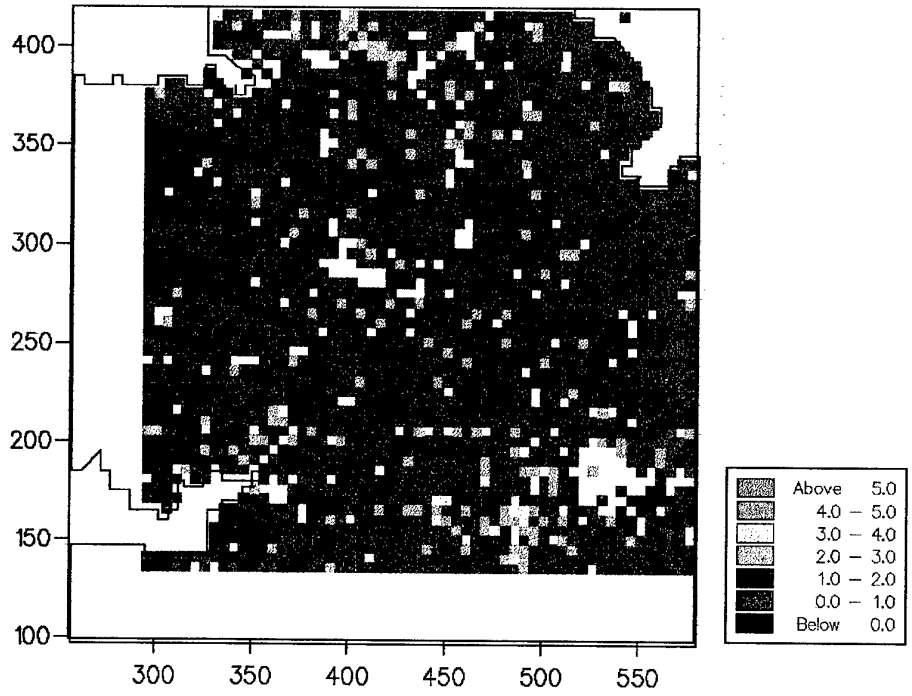


Figure 10. Comparisons for estimates on a 40-km grid

a) pH – comparisons for kriging



b) pH – comparisons for wavelet analysis

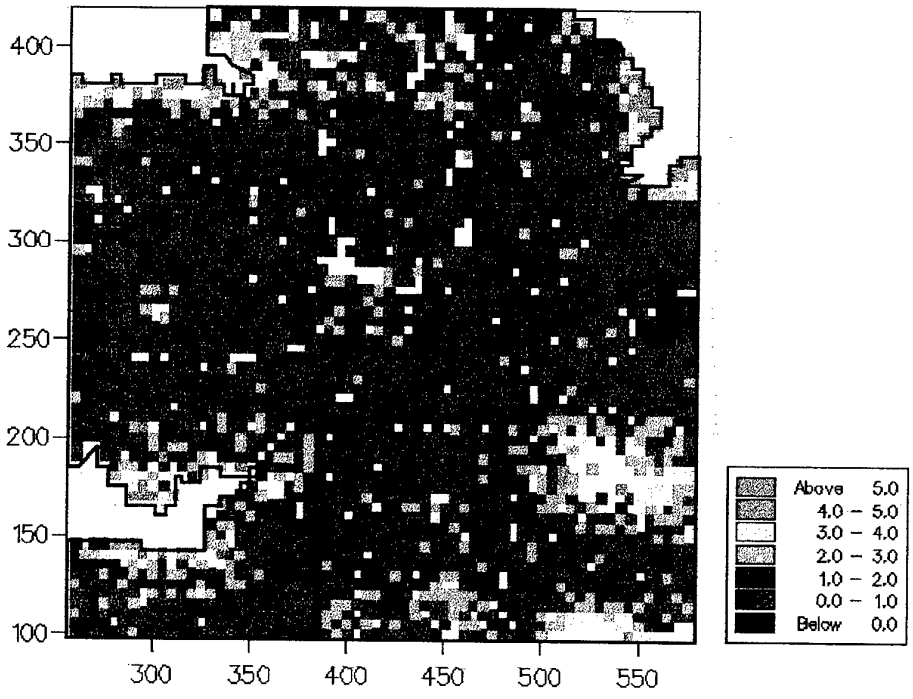
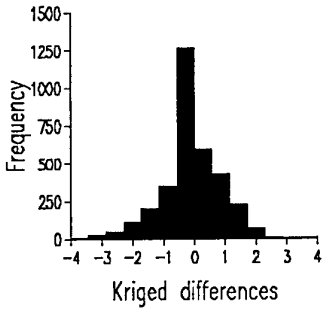
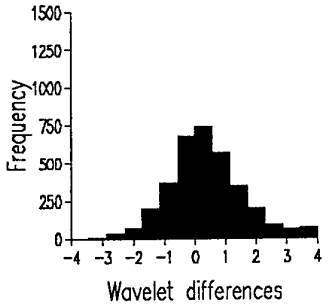


Figure 11. Histograms of the differences for pH from kriging and wavelet analysis.

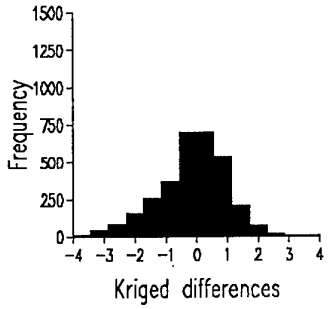
a) Kriged pH sampled at 1 in 4



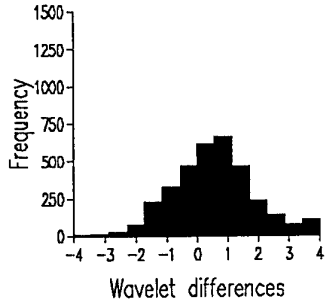
b) Wavelet pH sampled at 1 in 4



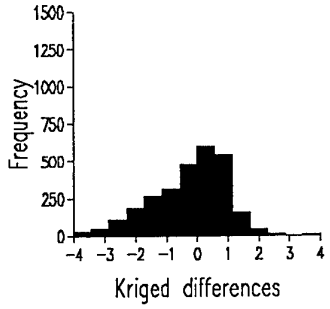
c) Kriged pH sampled at 1 in 16



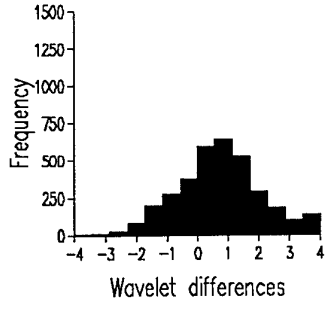
c) Wavelet pH sampled at 1 in 16



d) Kriged pH sampled at 1 in 64



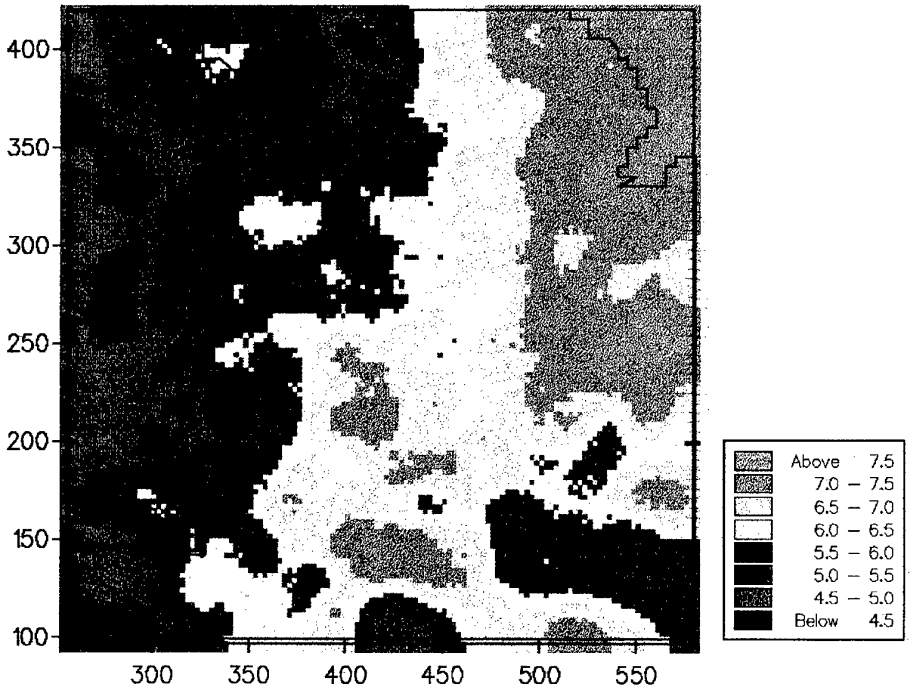
e) Wavelet pH sampled at 1 in 64



Results of factorial kriging and wavelet analysis for pH

Factorial kriging was applied to the data on the 5-km grid, but the equivalent analysis for wavelets was done on all of the subsamples.. The reasons for this were given in the previous final report. Figure 12 shows the long-range component from kriging analysis. The results for all of England and Wales are given at the end of the report, Figure 24.

Figure 12. Long-range estimates of pH from kriging analysis



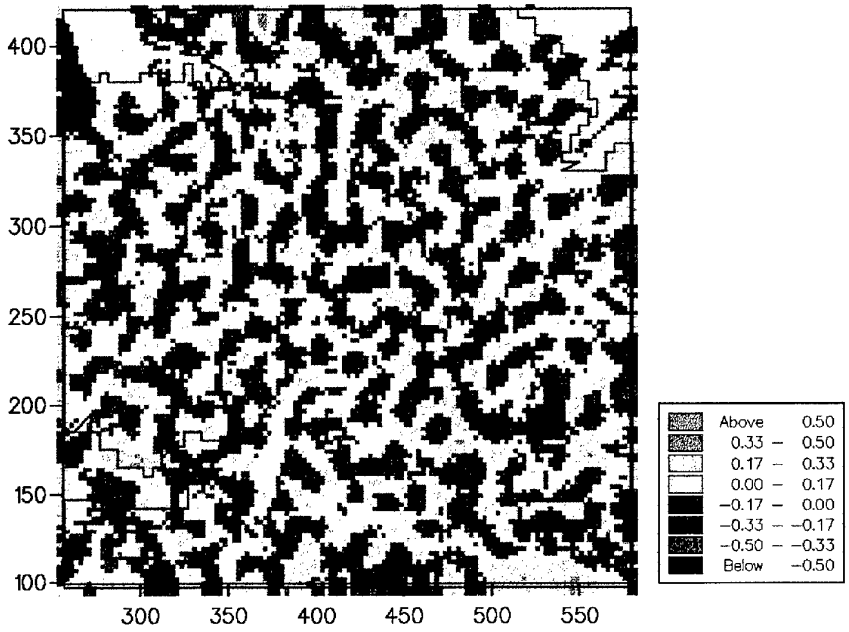
The map of the long-range estimates for pH is similar to the kriged estimates from the 20-km grid, they are not as similar to any of the low frequency wavelet predictions, Figures 5 to 7. The long-range variation, Figure 12 shows that the larger values are generally associated with the lowland areas and the limestone uplands. However, the western coastal areas have large values of pH which are most probably associated with the deposition of sodium ions by rain in these areas.

Figure 13 b, c and d shows the high frequency wavelet component for pH from the wavelet analysis. It is evident that the result for the 20-km grid is the closest. This reflects the same resolution for extracting the long-range component also. There are some similarities in the detail of the distributions, but there are also differences. In the future we shall examine the differences between these particular results to assess their relative performances in more detail. The high frequency component for the 40-km grid has not identified the relevant short-range component.

Again an interesting point emerges that we observed in the previous analysis of the SPOT data. The level of resolution at which the wavelet analysis has identified the long- and short-range components of the variation is related to the short-range parameter of the variogram. We can now suggest more forcibly that for a multiresolution analysis using wavelets the best approach is to compute the variogram first.

Figure 13. Short-range variation of pH

a) Short-range component of pH from kriging analysis on the 5-km grid



b) High frequency wavelet coefficient of pH from data on the 10-km grid

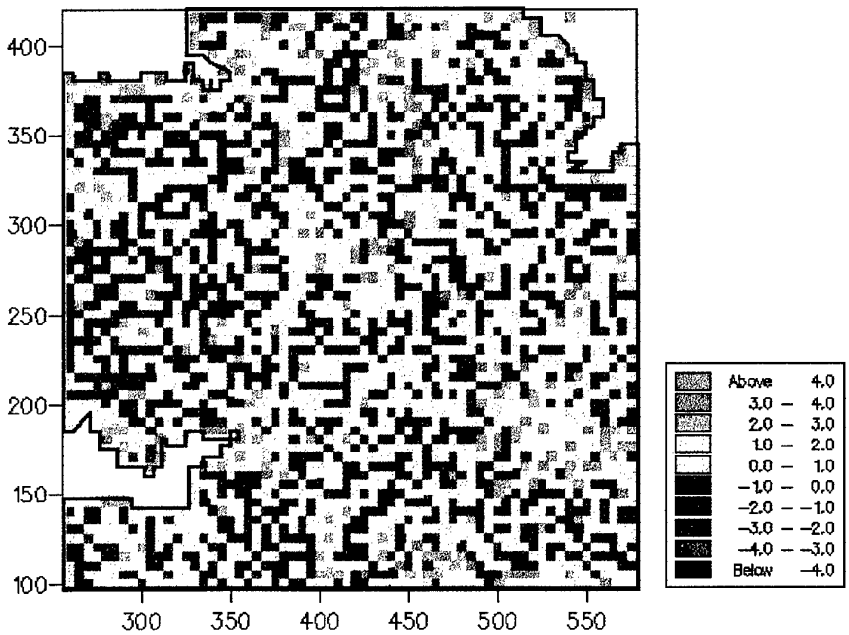
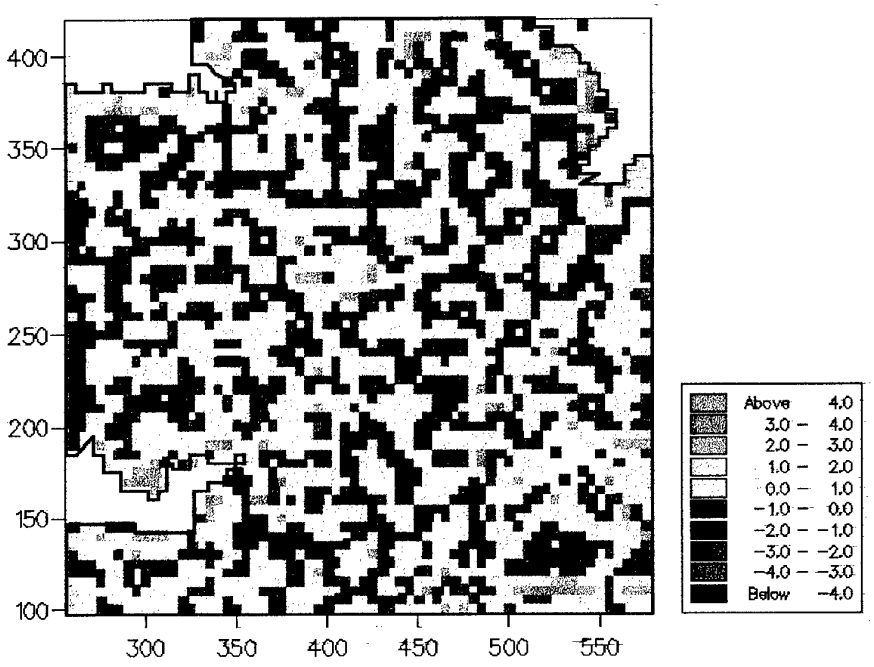
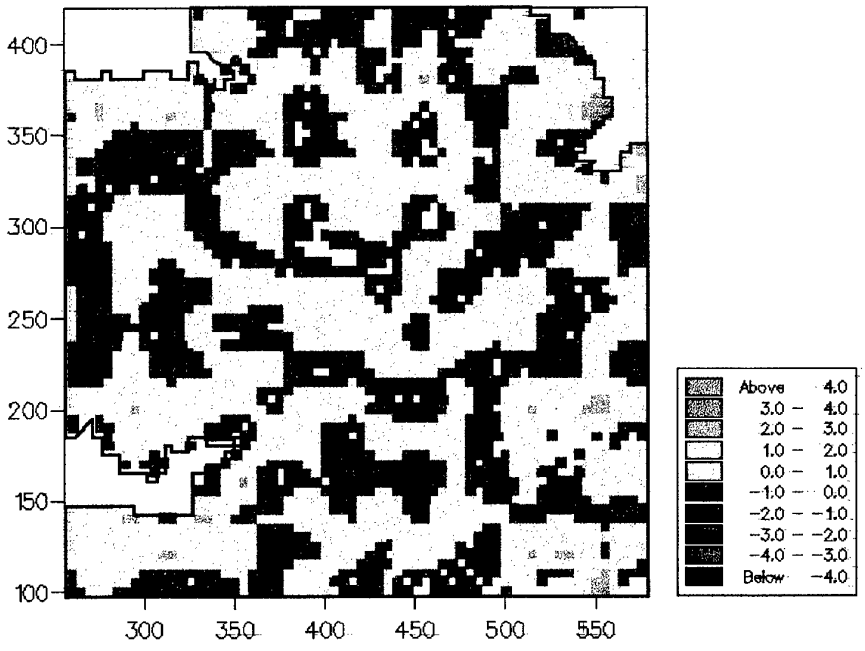


Figure 14. Short-range variation of pH.

a) High frequency wavelet coefficient of pH from data on the 20-km grid



b) High frequency wavelet coefficient of pH from data on the 40-km grid



RESULTS FOR ZINC

Figure 2 a and b show the original values for total $\log_{10}$Zn. There are many areas with large concentrations, and the most extensive of these follows the Jurassic clay band from SW to NE approximately. There are other areas trending N to S from the Midlands of England to the north. These seem to be associated with the Carboniferous shales and sandstones, as do the areas in central Wales. The large values around Avonmouth (SW) are associated with the smelting industry there.

For Zn the common logarithms were analysed and the values back-transformed for mapping as described above. Figure 15 a and b shows the maps of the predictions using ordinary punctual kriging and the low frequency wavelet coefficients for data on the 10-km grid. The results are similar. The spotty appearance of the kriged map arises from the fact that kriging restores the data at the sampling points with non error. The overall result shows that kriging smooths more than the wavelet analysis. Nevertheless the maps are similar to those for the original data, Figure 2 b.

The pattern of variation in the estimates from the 20-km grid for both analyses is also preserved well, Figure 16 a and b. The degradation in detail is clear, but the large-scale pattern is still evident. Again the results for both methods of analysis are similar – more so than for pH.

Figure 17 a and b shows the results for the 40-km grid. The results from the wavelet analysis, although showing a loss of detail, still show a similar pattern of variation, Figure 17 b, to that of Figure 16 b. The kriged results so not show such a good resemblance to the original pattern of variation. Note particularly the loss of accuracy in the north western part of the country.

These results again accord with the findings for pH and for the analysis of the SPOT image. When the number of samples is few and the distance between them large kriging restores the data less well than the wavelet analysis. This effect is supported by the maps of the differences, Figures 18 to 20 and of the histograms, Figure 21.

Figures 18 to 20 show the maps of the absolute differences between the predictions from ordinary punctual kriging and the low frequency wavelet transform, and the original values at the sampling sites of the 5-km grid. Figure 18 a and b shows the comparisons, for predictions based on the 10-km sampling grid. Figure 18 a for the kriged comparisons is a more spotty map than the one from the wavelet analysis: the sampling points are evident as the blue pixels where there is no error. For the wavelet analysis for this sampling grid there are fewer zero or small errors than for kriging, This is confirmed by the histograms of the differences, Figure 21 a and b. The same negative skew in the errors is evident for Zn as for pH. Kriging does not appear to have performed quite as well for Zn for the 20-km grid as the wavelet analysis in terms of the small errors, Figure 21 c and d. The maps of the differences, Figure 19 a and b do not show this as clearly. Both methods appear to have performed similarly from these two maps. The slight negative skewness in this histogram again suggests that there is some bias in the predictions. The comparisons for the predictions from the 40-km grid suggest that there is less difference between the wavelet analysis and kriging than the maps of the estimates suggested there would be, Figure 17. The histograms, Figure 21 d and e confirm this, although direct comparison is not possible because kriging has not gone to the edges of the area for the reasons given before.

Summary

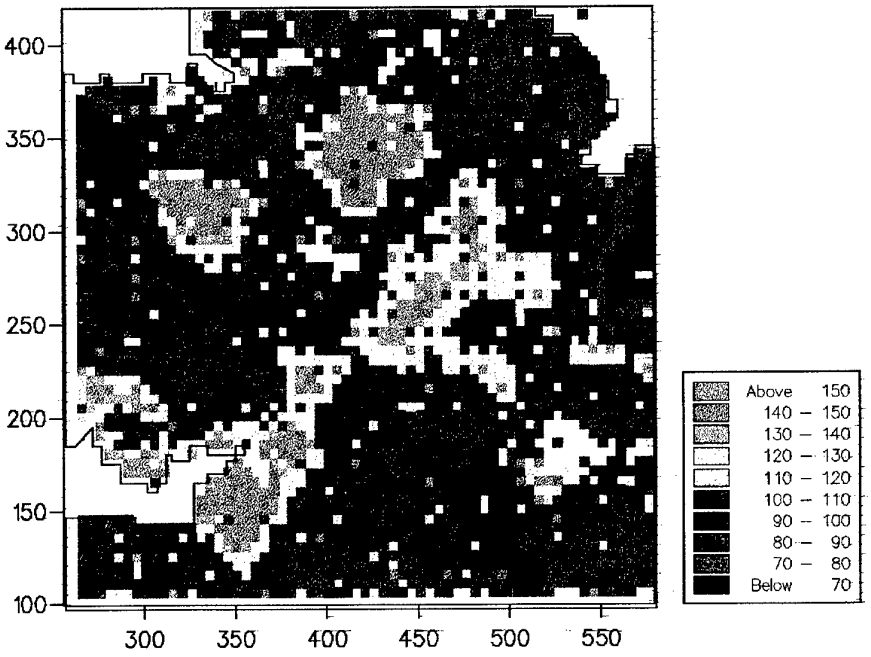
These results are interesting when compared with the analysis of the SPOT data. The NSI data for zinc again do not appear not contain locally non-stationary data as for pH. This explains the somewhat better performance of kriging for the 10-km grid.

The zinc values were skewed and this means that the variances when computing the variogram are unstable. The values were transformed to common logarithms, $\log_{10}$Zn,

and kriging was performed on the logarithms and these values were back-transformed afterwards so that the values could be shown on their original measurement scale as for the wavelet analysis. Again this is clearly involves additional effort over and above the wavelet analysis, which does not require non-normal distributions to be transformed. This has an additional advantage because the transformation causes additional smoothing of the predictions. This does not seem to be particularly evident from the results given here.

Figure 15. Predictions of Zn at a 5-km interval from data on a 10-km grid

a) Zn - kriged estimates from data on the 10-km grid



b) Zn – low frequency wavelet coefficients from data on the 10-km grid

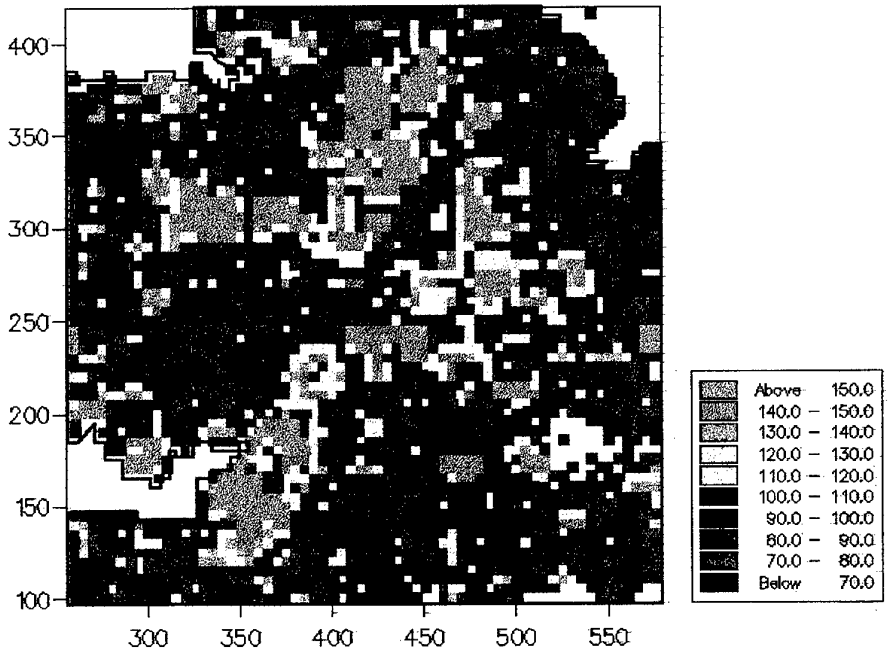
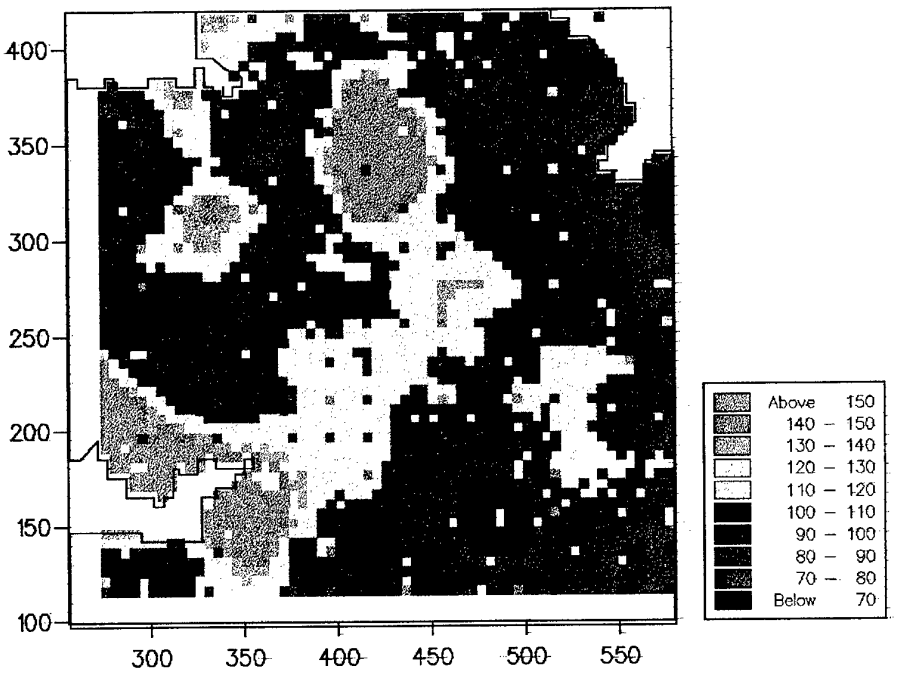


Figure 16. Predictions of Zn at a 5-km interval from data on a 20-km grid

a) Zn – kriged estimates from data on the 20-km grid



b) Zn – low frequency wavelet coefficients from data on the 20-km grid

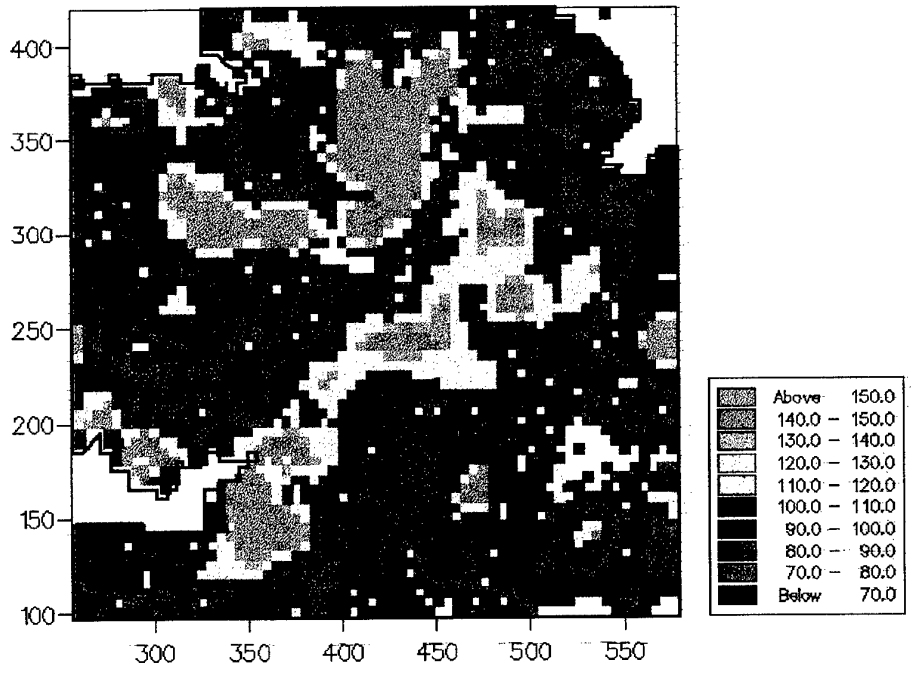
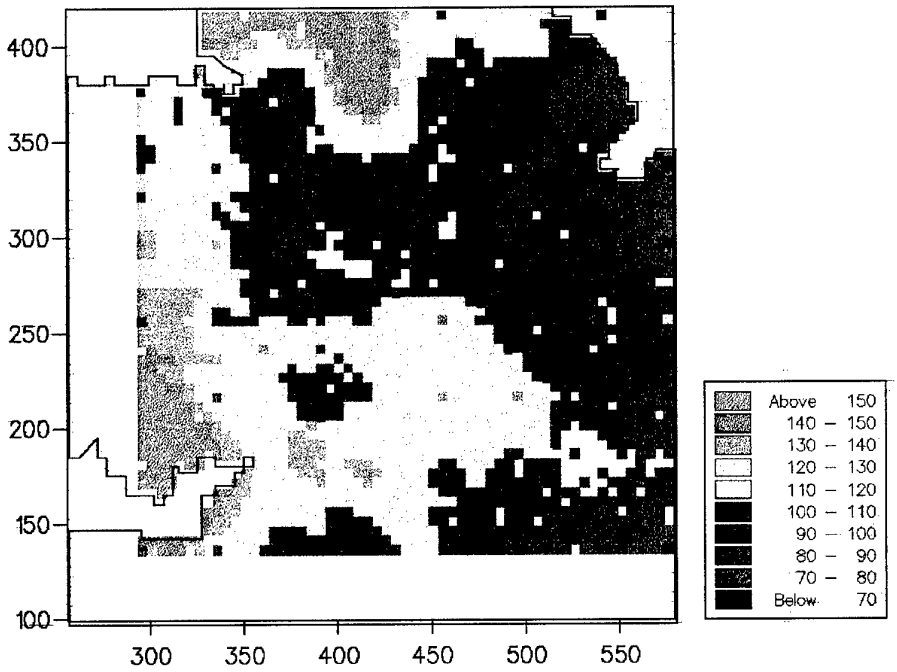


Figure 17. Predictions of Zn at a 5-km interval from data on a 40-km grid

a) Zn – kriged estimates from data on the 40-km grid



b) Zn – low frequency wavelet coefficients from data on the 40-km grid

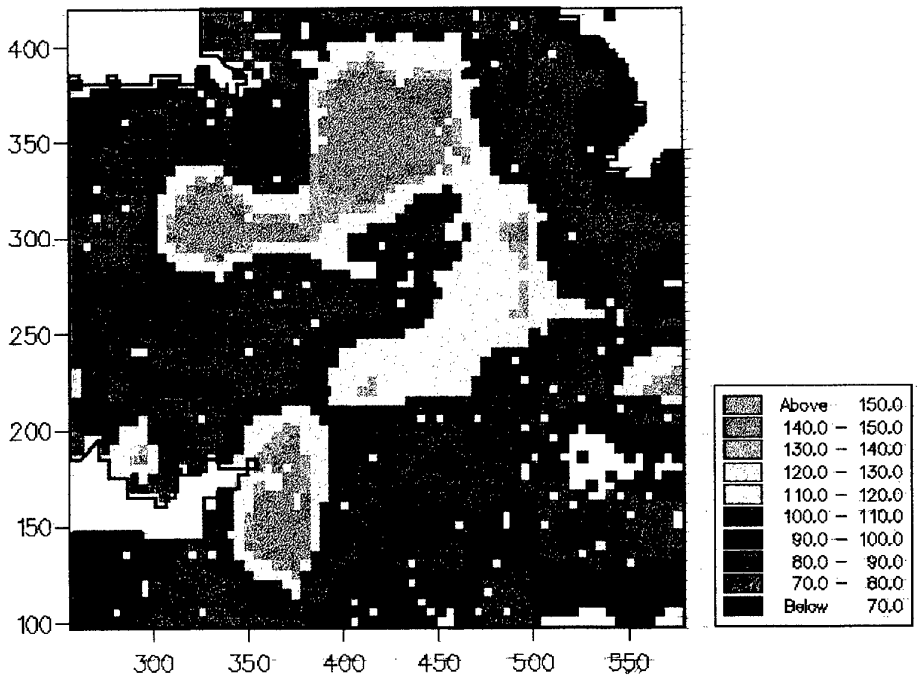
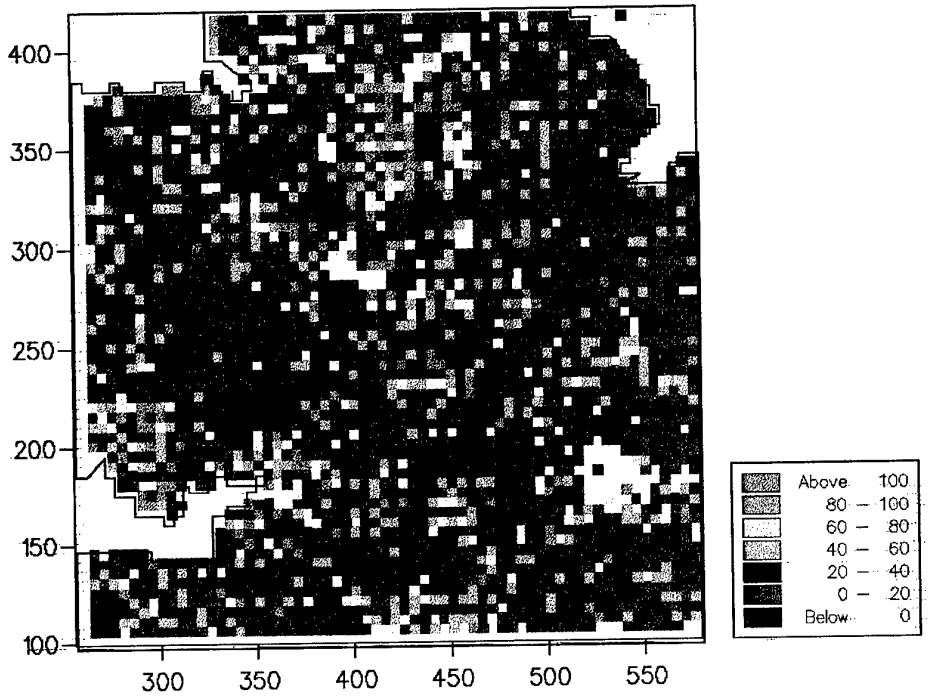


Figure 18. Comparisons for estimates on a 10-km grid

a) Zn – comparisons for kriging



b) Zn – comparisons for wavelet analysis

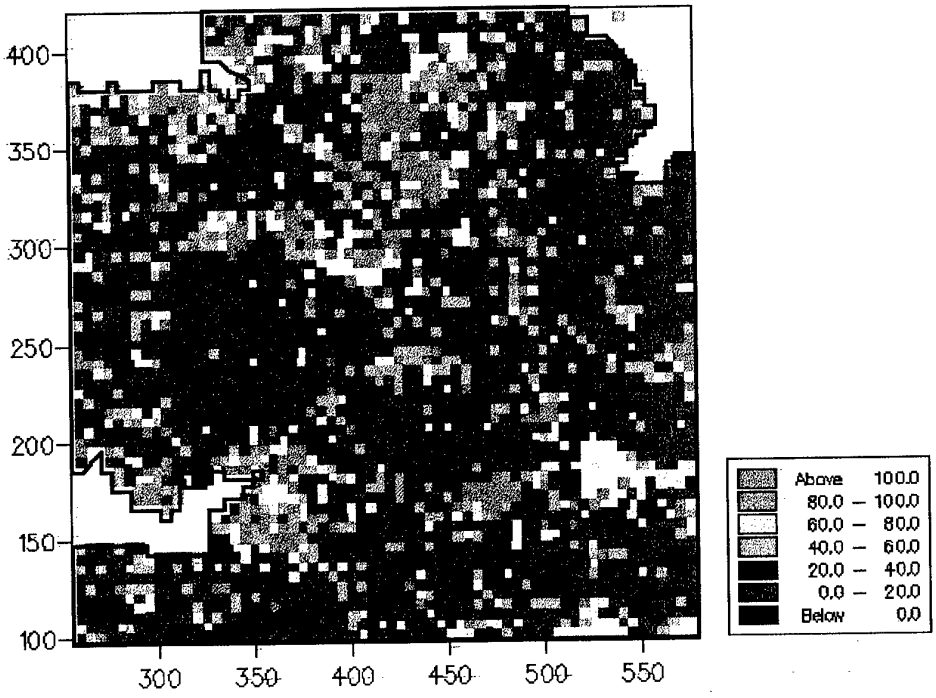
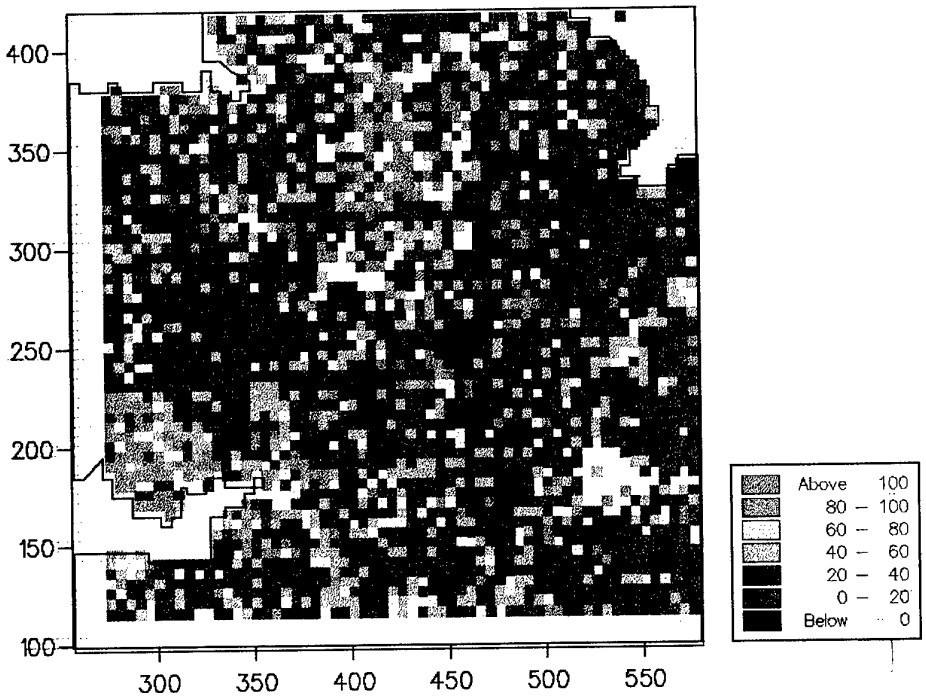


Figure 19. Comparisons for estimates on a 20-km grid

a) Zn – comparisons for kriging



b) Zn – comparisons for wavelet analysis

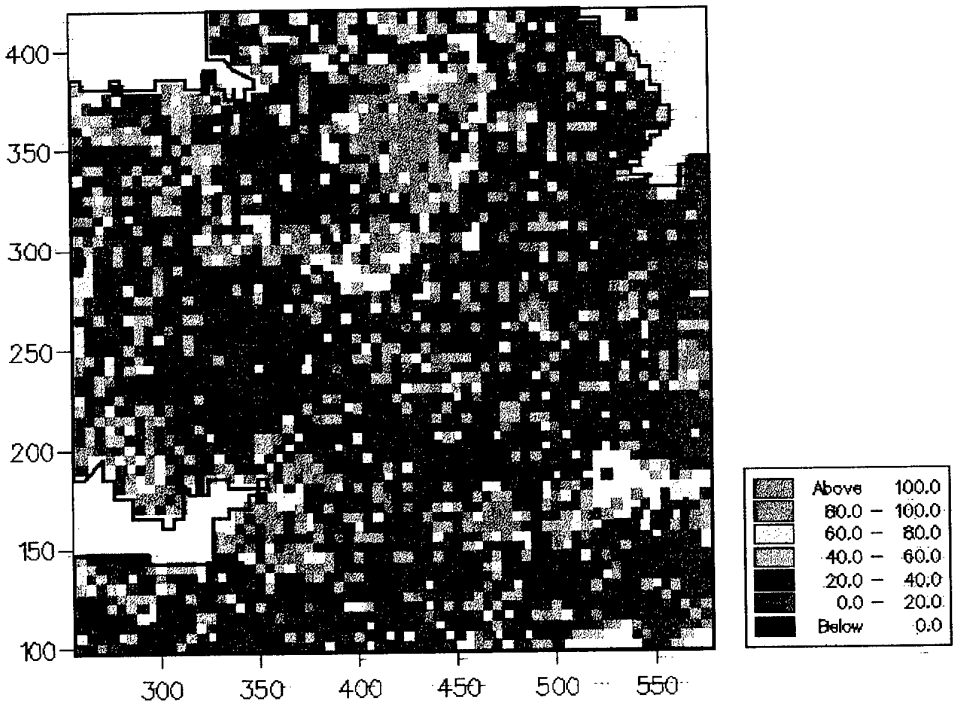
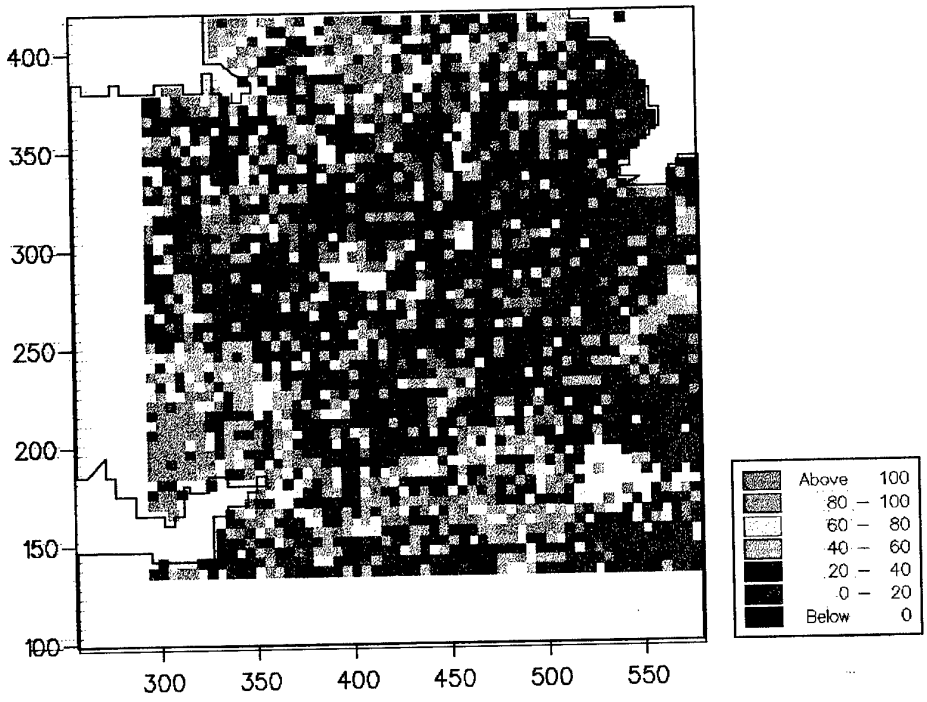


Figure 20. Comparisons for estimates on a 40-km grid

a) Zn – comparisons for kriging



b) Zn – comparisons for wavelet analysis

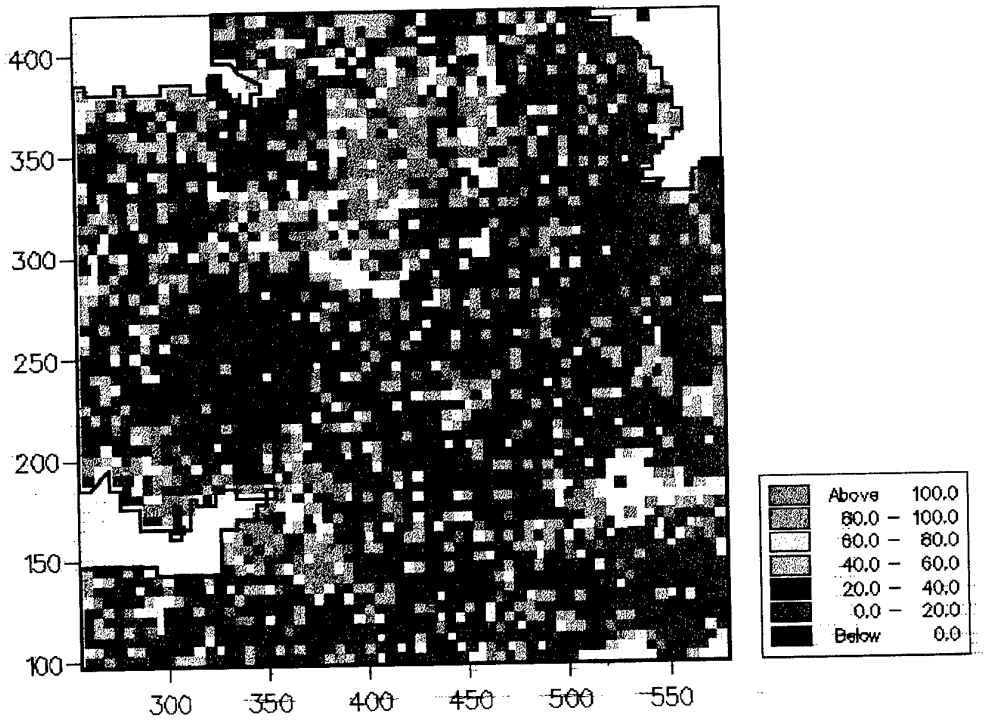
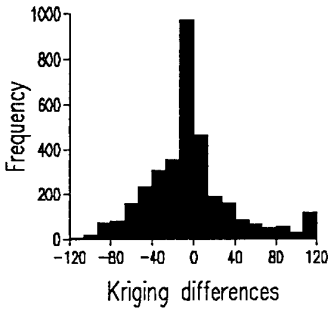
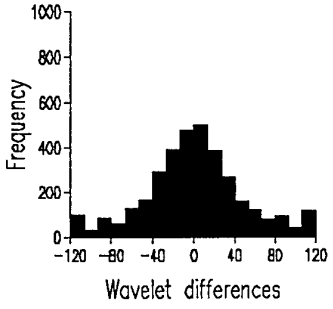


Figure 21. Histograms of Zinc

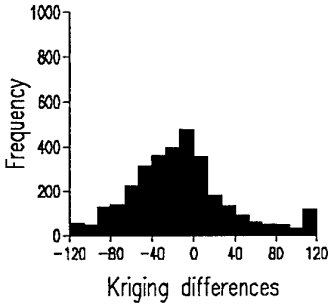
a) Kriged Zn sampled at 1 in 4



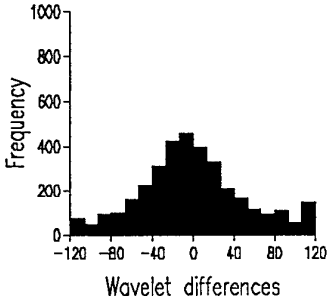
b) Wavelet Zn sampled at 1 in 4



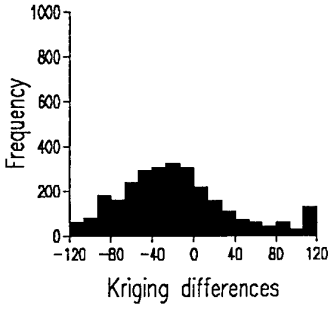
c) Kriged Zn sampled at 1 in 16



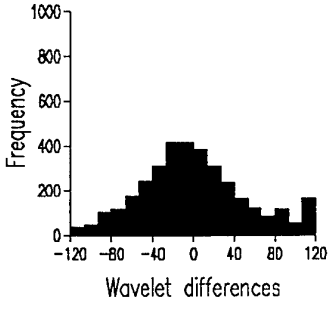
c) Wavelet Zn sampled at 1 in 16



d) Kriged Zn sampled at 1 in 64



e) Wavelet Zn sampled at 1 in 64



Results of factorial kriging and wavelet analysis for pH

Factorial kriging was applied to the data on the 5-km grid, but the equivalent analysis for wavelets was done on all of the subsamples.. The reasons for this were given in the previous final report. Figure 22 shows the long-range component from kriging analysis on the logarithmic scale. These results were not back-transformed because of the way in which these estimates are derived. The long-range kriged estimates for $\log_{10}$ Zn show that the largest values occur near to the Avonmouth smelter in the west of England, and another area in Derbyshire. There are large values associated with the Jurassic clays trending from SW to NE, to the Carboniferous limestone in Derbyshire, Carboniferous shales in the NE and Ordovician rocks in Wales. This distribution has some similarities with that for Cr. The values for $\log_{10}$ Zn for all of England Wales are shown in the end. These results show the closest relations with the 20-km grid wavelet analysis, Figure 16 b, even though the colour scales appear somewhat different because Figure 22 is for logarithms.

Figure 22. Long-range component from factorial kriging for Zinc.

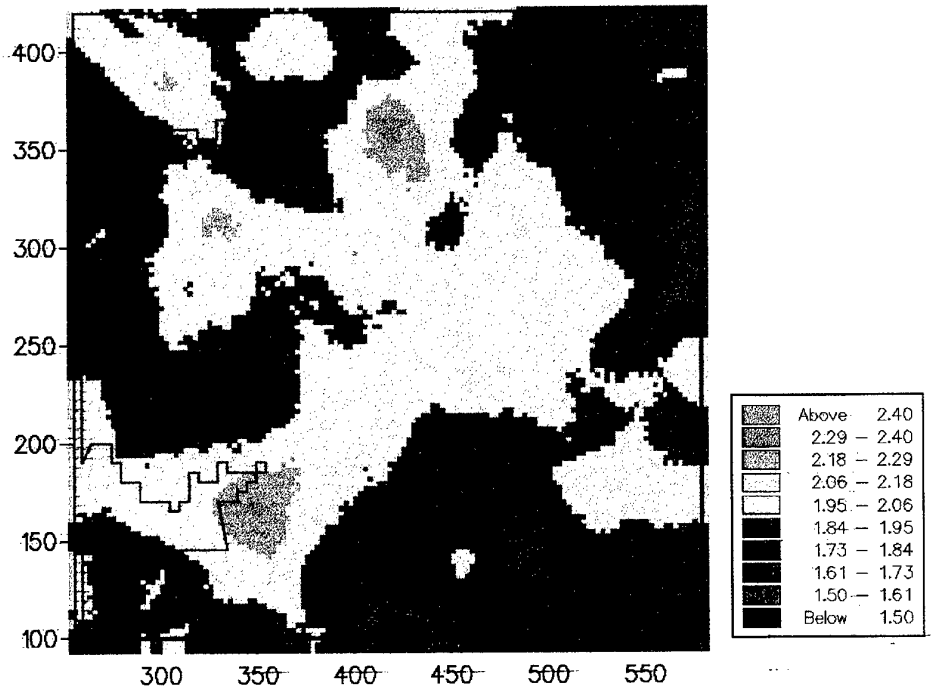
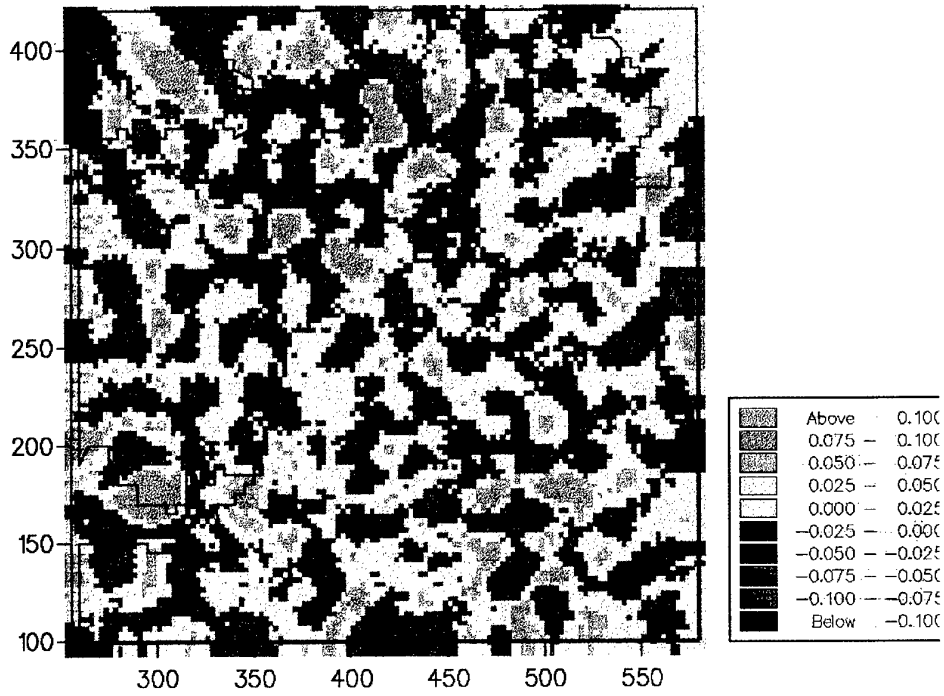
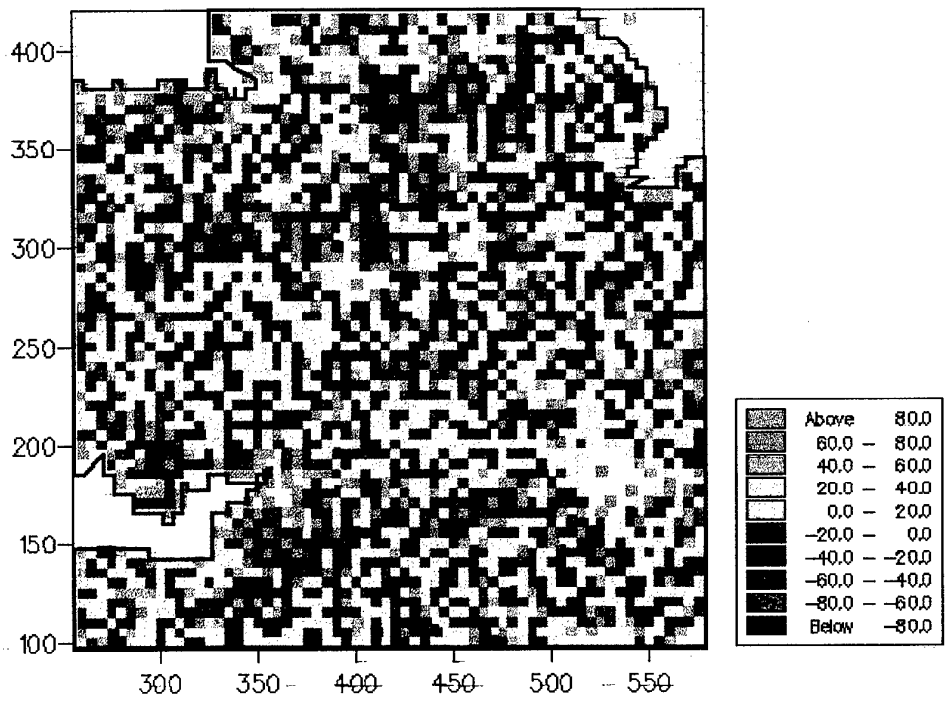


Figure 23 Short-range spatial component of Zinc

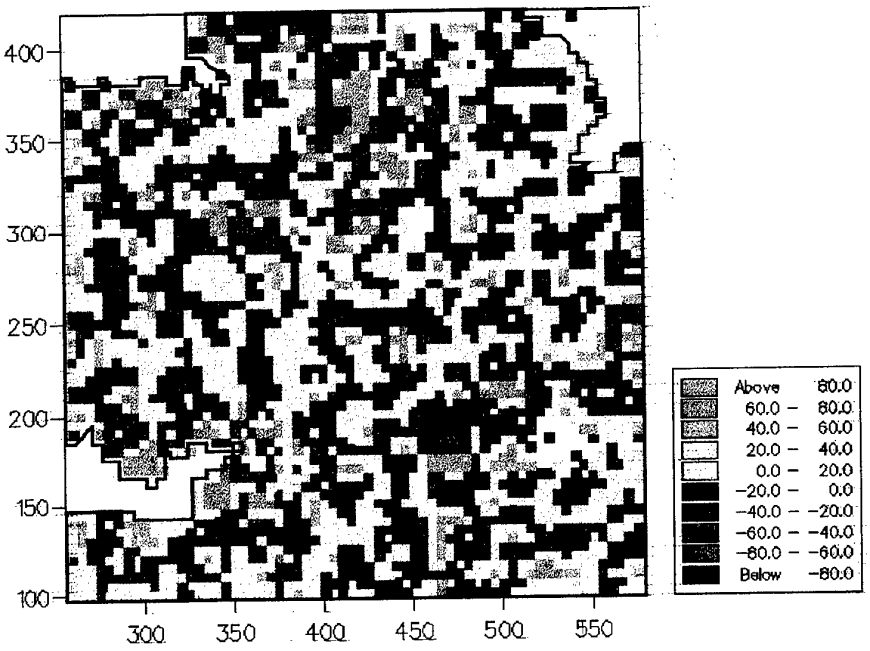
a) Zn – short-range component from factorial kriging



b) Zn – high frequency wavelet coefficient for data on the 10-km grid



c) Zn – high frequency wavelet coefficient for data on the 20-km grid



d) Zn – high frequency wavelet coefficient for data on the 40-km grid

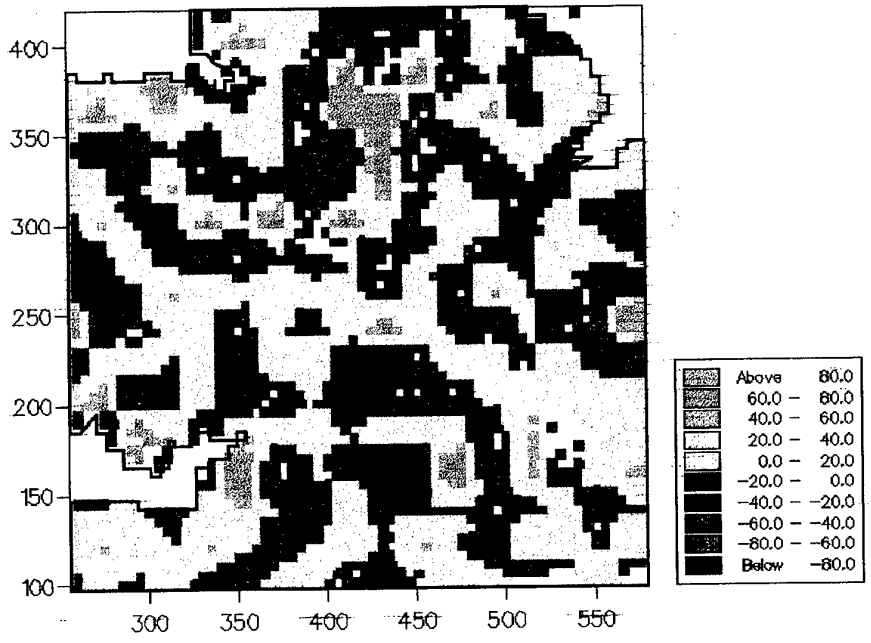


Figure 23 a shows the short-range component of the variation from kriging analysis, and the maps for the whole of England and Wales for this analysis are given at the end of the report, Figure 25. The short-range component was investigated previously, it has a strong similarity with the map of the short-range component for Cr (not shown). These distributions also show a relation with the small scale drainage basins and local changes in rock and soil types.

Figure 23 b, c and d shows the high frequency wavelet component for Zn from the wavelet analysis. It is evident that the result for the 20-km grid is the closest to that for the short-range component from kriging analysis. This reflects the same resolution for extracting the long-range component also. There are some similarities in the detail of the distributions, but there are also differences. In the future we shall examine the differences between these particular results to assess their relative performances in more detail. The high frequency component for the 40-km grid has also identified some of the relevant short-range component.

Again an interesting point emerges that we observe above is that the level of resolution at which the wavelet analysis has identified the long- and short-range components of the variation is related to the short-range parameter of the variogram.

References

Avery, B.W. 1987. *Soil Survey Methods: a review.* Technical Monograph No. 18. Soil Survey and Land Research Centre, Silsoe.

Hodgson, J.M. 1974. *Soil Survey Field Handbook*, 1st edn. Technical Monograph No. 5. Soil Survey of England and Wales, Harpenden.

Loveland, P.J. 1990. *The National Soil Inventory of England and Wales.* In: *Element Concentration Cadasters in Ecosystems* (eds H. Lieth and B. Markert), pp. 73-80. VCH, Weinheim, Germany.

MAFF. 1986. *The Analysis of Agricultural Materials.* Reference Book 427. HMSO, London.

McGrath, S.P. & Cunliffe, C.H. 1985. A simplified method for the extraction of the metals Fe, Zn, Cu, Ni, Cd, Pb, Cr, Co and Mn from soils and sewage sludges. *Journal of the Science of Food and Agriculture* **36**, 794-798.

McGrath, S.P. & Loveland, P.J. 1992. *The Soil Geochemical Atlas of England and Wales.* Blackie Academic and Professional, Glasgow.

Oliver, M. A., Webster, R, and Slocum, K. 2000. Filtering SPOT imagery by kriging analysis. *International Journal of Remote Sensing*, **21**, 735-752.

Waters, D.F. & Sweetman, I.C. 1955. The Rukuhia Soil Grinder. *Soil Science* **79**, 411-413.

Acknowledgement

The Natiional Soil Iventory of England and Wales was funded by the Ministry of Agriculture, Fisheries and Food of London.

Figure 24 . pH – results of factorial kriging: short- and long-range components.

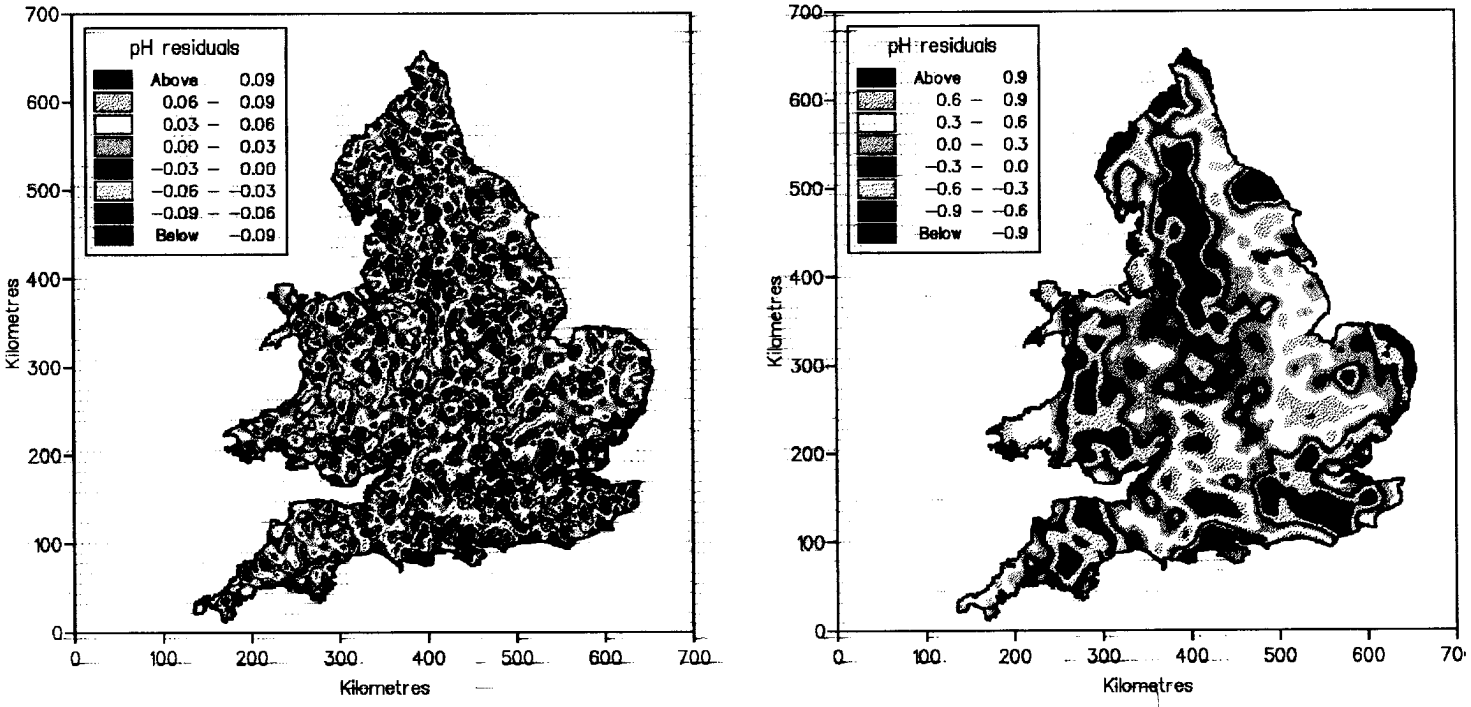
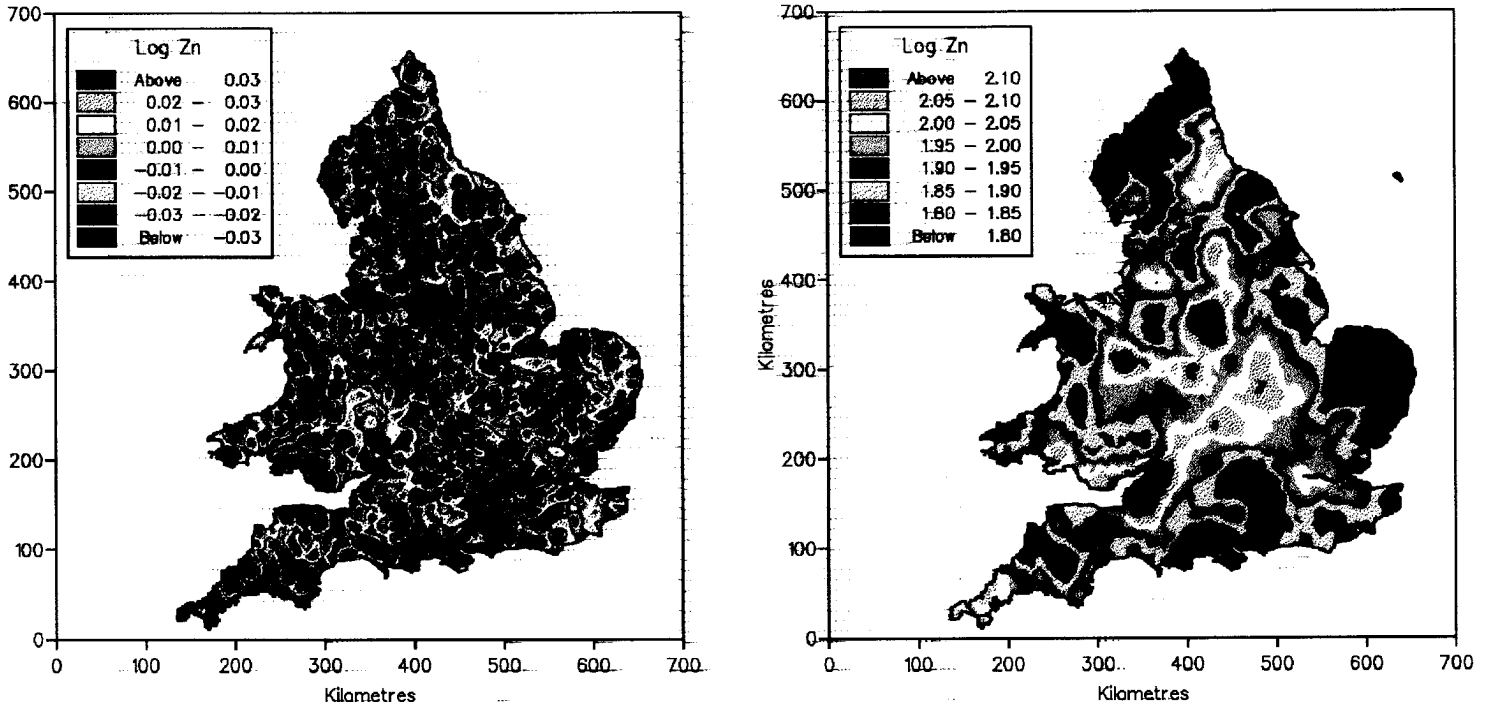



Figure 25. Zinc – results of factorial kriging: short- and long-range components.